THE
PHYSICAL REVIEW
A Journal of Experimental and Theoretical Physics Established by E. L. Nichols in 1893
VOL. 58, NO. 4 AUGUST 15, 1940 SECOND SERIES

A Study of Time-Variations in Cosmic-Ray Intensity at High Altitudes*

WILLIAM P. JESSE
Ryerson Laboratory, University of Chicago, Chicago, Illinois
(Received June 17, 1940)

Twelve balloon flights with a recording cosmic-ray ionization chamber apparatus have been carried out at Chicago between October, 1938 and November, 1939. The results show marked changes in cosmic-ray intensity during this period, the maximum variation observed being about fourteen percent for the peak value of the ionization pressure curve. The changes follow in general the "world-wide" variations of cosmic rays observed at ground stations but when the high altitude results are somewhat arbitrarily corrected for such "world-wide" variations certain residual changes remain. These residual variations show a maximum in the early spring of 1939, then a sudden drop of about eight percent and lower values during the early summer with increasing values again in the autumn. Whether this represents a true seasonal effect can be determined only by further experiments. Though there is some indication that the variations are related to magnetic changes, the agreement with the theory of Vallarta and Godart is not quantitatively good.

DURING the past eighteen months a series of flights with unmanned balloons carrying cosmic-ray apparatus has been carried out at Chicago to determine whether any change with time could be observed in the intensity of cosmic rays at very high altitudes.

APPARATUS

The apparatus was of the recording type and had a mass of from 3000 to 4000 grams, the exact value depending largely upon the type of battery used for the electrometer quadrants. The ionization chamber method was used, the chamber consisting of a sphere 15 cm in diameter with 0.05-cm steel walls and filled to 10 atmospheres of argon. The most important part of the apparatus was an electrometer designed by Dr. Elmer Dershem of this laboratory. This electrometer was fitted with an especially rugged quartz suspension and this was meticulously balanced to give the same deflection with the electrometer oriented in any position. The instrument was thus rendered very insensitive to mechanical shock. A beam of light was reflected from the mirror on this instrument and the image of the light source traversed the length of a slit behind which moved a drum carrying a strip of bromide paper. The insulated system was grounded every 3½ minutes and the electrometer voltage sensitivity was recorded on every fifth grounding.

The pressure device was of the metal bellows type, the bellows being very flexible and the steel spring very stiff. Extensive tests in a vacuum chamber on the ground showed that the barometer element would give pressure readings reproducible to within one millimeter of mercury.

* A preliminary report on these experiments was made at the Symposium on Cosmic Rays in June, 1939 and was published in Rev. Mod. Phys. 11, 167 (1939).

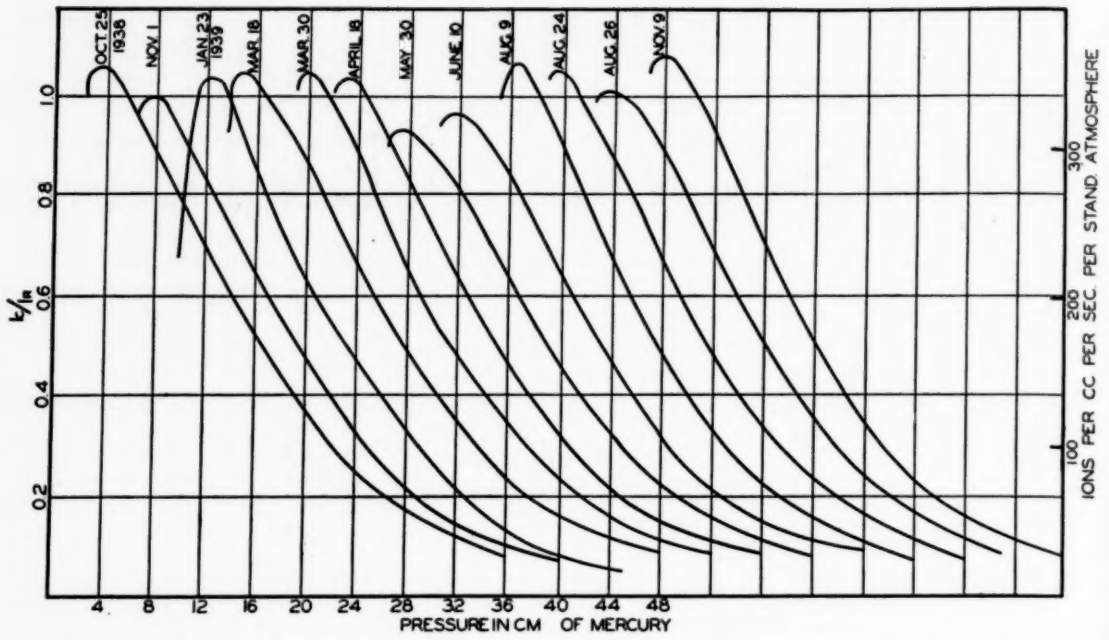


FIG. 1. Ionization due to cosmic rays relative to radium standard (I_c/I_r) as a function of pressure. Each curve has been arbitrarily shifted to the right with respect to the preceding by an amount equal to 4 cm pressure.

Slight hysteresis effects could be observed in the first barometer elements used but such effects were entirely eliminated in later instruments. No effect of temperature upon the reading of the instrument could be observed.

The temperature variation of the apparatus within the gondola was kept within reasonable limits by the usual device of a Cellophane covering with strips of silver paper to reflect a portion of the sunlight. Such a device maintained a temperature constant to within 15°C. The temperature of the apparatus was recorded by means of a bimetallic strip carrying a small mirror.

Before each flight the set was calibrated in terms of a radium standard at a standard distance and with a standard geometrical orientation. All cosmic-ray values were reckoned relative to this standard calibration. For purely relative cosmic-ray measurements there was thus avoided the necessity of knowing with great accuracy the absolute value of the electrical capacity, the timing interval or the density of the gas in the ionization chamber, although all these values were measured. An experiment to determine the effect of temperature upon the ionization chamber measurements was made by varying the temperature of the whole apparatus with the radium source in the standard position. An increase of temperature of 15°C. caused no detectible change in the length of the electrometer trace registered on the bromide paper. Thus any possible change under these conditions was less than 1 percent of the total ionization recorded.

EXPERIMENTAL RESULTS

Between October 25, 1938 and November 9, 1939 twelve flights were carried out. Ten of these were made with the same apparatus which was each time recovered with a recorded trace which was usable. The other two flights were made with a second similar apparatus. The results are shown in Fig. 1 where ionization relative to the radium standard calibration (I_c/I_r), is plotted against pressure in centimeters of mercury. To avoid confusion on the diagram, each curve is arbitrarily shifted with regard to the preceding curve along the abscissa axis by a step of 4 cm pressure. A number of determinations give for the ionization produced in the chamber by the radium in the standard position a mean value of 465 ion pairs per cc per standard atmosphere of argon. If one takes the relative ionization of air to argon at one atmosphere under standard con-

ditions to be 0.72,[1] then one obtains the value 335 for the number of pairs of ions per cc per sec. in standard air produced by the gamma-rays from the radium. If the ordinate ratios (I_c/I_r), for the curves in Fig. 1 are multiplied by this value one obtains absolute values for the ionization produced by cosmic rays. An ordinate scale for such absolute values is shown to the right of the figure.

The curves in Fig. 1 are similar in shape to ionization-pressure curves obtained by other investigators. It will be seen that in each flight sufficient altitude has been attained to outline definitely the peak of the curve. The height of the curves varies markedly from flight to flight, the maximum variation in the peak values being about fourteen percent. The corresponding change in the integrated area under the curves is about ten percent. This represents the relative change in the total energy of cosmic rays entering the earth's atmosphere. The variations in the ionization chamber readings at different pressures are shown more clearly in Fig. 2 where these values for different pressures are plotted as a function of time. For each pressure shown a mean of the corresponding ionization chamber readings for the twelve flights has been determined and the ordinate values in Fig. 2 are reckoned for each pressure in percent of this mean value. The ordinate scale in the lower left corner of the figure applies to each of the eight curves.

[1] E. F. Cox, Phys. Rev. **45**, 503 (1934).

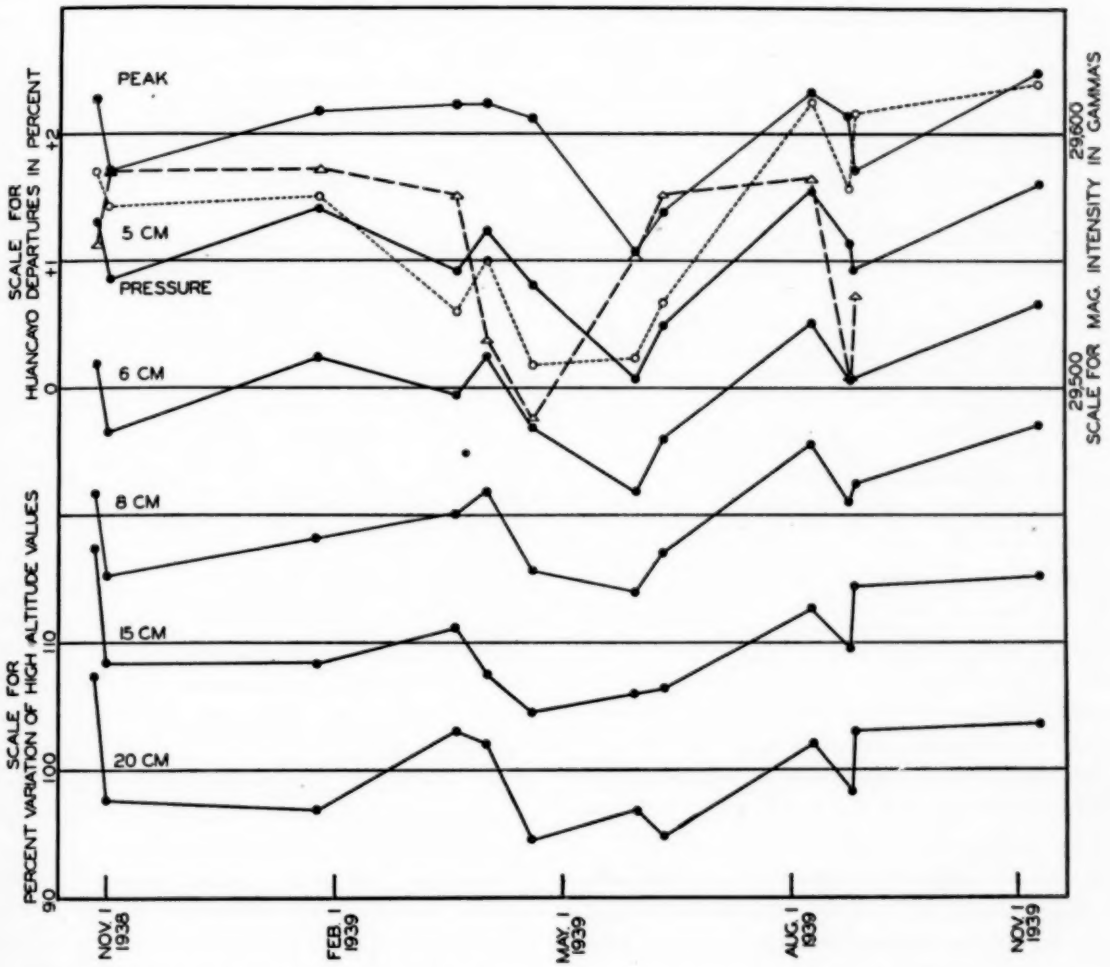


FIG. 2. Plots against time for: ●, High altitude values in percent of mean value for eight different pressures; ○, departures from mean value in percent for cosmic-ray meter at Huancayo, Peru; △, horizontal component of magnetic intensity in gammas at Huancayo Station.

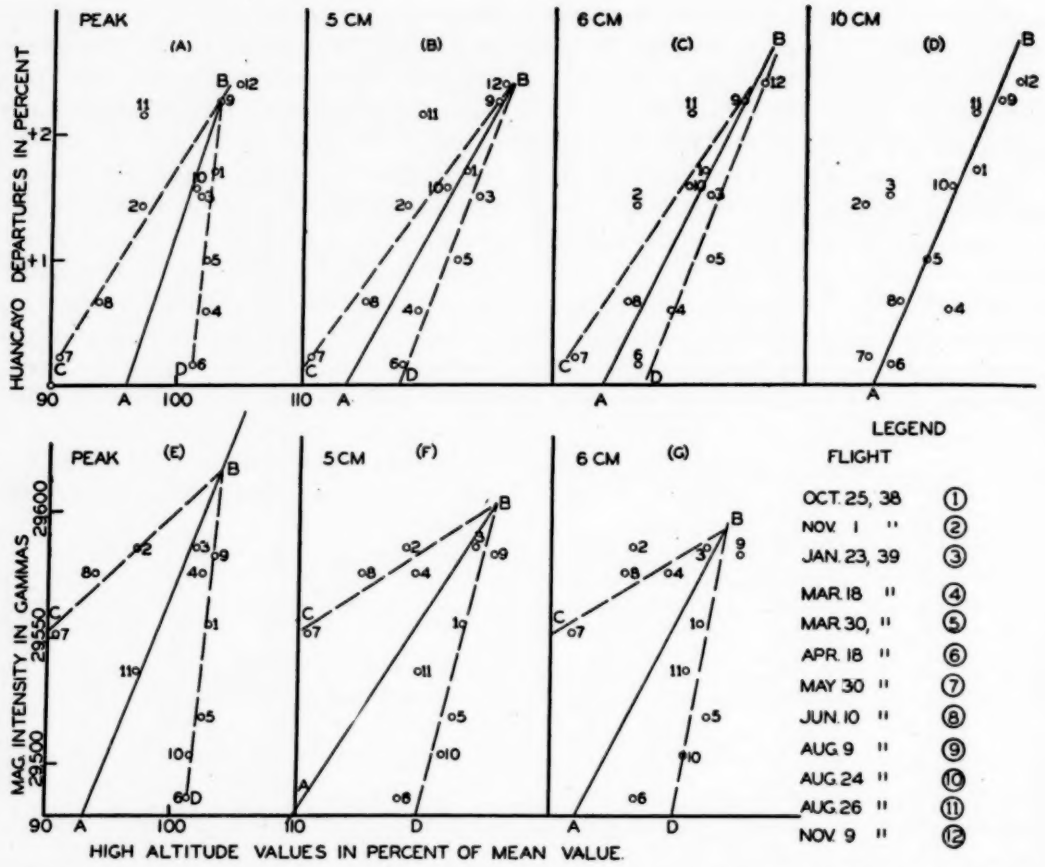


FIG. 3. In top row of diagrams, Huancayo departures are plotted against high altitude cosmic-ray values. In lower row, Huancayo magnetic intensities are plotted against high altitude values. (Lines *CB* and *DB* are *not* regression lines.)

DISCUSSION OF RESULTS

The first evidence that there might be a time change in the intensity of cosmic rays at high altitudes was obtained by investigators[2] in the laboratory of Regener. A flight in December, 1934 showed cosmic-ray intensities higher by six percent than those given by a flight a month earlier. This change was considered greater than the experimental error but was attributed to the activity of a nova. Also during the progress of the present work, papers have appeared by Millikan and Neher[3] in which they have observed similar time variations in cosmic-ray intensity at high altitudes. They observed a difference in intensity of about 9 percent between flights made at Omaha, Nebraska in September, 1937 and in December, 1938 and also a difference of about 5 percent between flights made at Saskatoon in August, 1937 and those at Bismark, North Dakota, in July, 1938. The change in intensity in each of these two cases was in the same direction as the intensity change during the same period of the "world-wide" variations which Forbush[4] has reported. The authors were thus led to suggest that the high altitude variations follow, at least to some extent, such "world-wide" variations observed on the ground. This prediction seems to be borne out in the more extensive data obtained in the present experiment.

The dotted line in Fig. 2 represents the variations observed in cosmic-ray intensity with a Model C meter[5] at the mountain station of the

[2] R. Auer, Zeits. f. Physik **111**, 559 (1939).
[3] R. A. Millikan and H. V. Neher, Phys. Rev. **56**, 491 (1939).
[4] S. E. Forbush, Phys. Rev. **54**, 975 (1938).
[5] A. H. Compton, E. O. Wollan and R. D. Bennett, Rev. Sci. Inst. **5**, 415 (1934).

Carnegie Institution at Huancayo, Peru (geomag. lat. 0.6 S; altitude, 3350 m). At this latitude the seasonal variation is negligible so the readings give directly the "world-wide" variations which Forbush and others have observed and which presumably follow changes in the magnetic field surrounding the earth. Each point represents the mean departure for a three-day period from balance of the instrument in percent of the mean cosmic-ray intensity, an average being taken of the daily mean for the day of the flight, the day before and the day after. Should one take only the mean for the day of the flight no important change would be made in the values given. It will be seen that there is a definite similarity between this curve and that for the high altitude values. The main features of the Huancayo curve are reproduced in the high altitude curves even for pressure of 15 cm and of 20 cm of mercury where the errors of experiment are much larger than for the lower pressures. It should be noted that the scale of the Huancayo values is ten times as large as that for the high altitude results so that the magnitude of the percent change in the latter is much greater.

In Fig. 2 is also plotted a curve showing variations in the horizontal intensity of the earth's magnetic field observed at the same Huancayo station. The readings are in gammas (1 gamma $=10^{-5}$ gauss) and represent also three-day means. The similarity between this curve and the high altitude curves is not as striking as for the Huancayo cosmic-ray values although the two curves present certain similar features. As has, however, been pointed out by Hess,[6] the magnetic changes measured on the surface of the earth do not necessarily present a perfect criterion by which to judge magnetic changes far out in space where the cosmic rays are most affected.

In the upper row of diagrams in Fig. 3 the high altitude values are shown plotted as abscissae against the corresponding Huancayo cosmic-ray values as ordinates for four different pressures. In the second row of diagrams the Huancayo magnetic values are plotted as ordinates. As will be seen, the experimental points do not lie very closely upon any one straight line which can be drawn through the diagram. This scatter of the points might, of course, be considered as caused by errors of experiment. This, however, would assume experimental errors far larger than it is believed exist in the data. Thus in Fig. 3(*A*) points 6 and 7 have about the same ordinate values but differ by more than ten percent in abscissa values. These points represent results from two of the best flight records and the probable error in these peak values is estimated to be not more than one percent. Moreover, in comparing diagrams 3(*A*), (*B*), (*C*), (*D*) where the experimental accuracy progressively decreases with increasing pressure, it may be seen that the scatter of the points from the median trace *AB* is worst in 3(*A*) where the accuracy is probably the best. The correlation coefficients for the four diagrams are, respectively, 0.49 $\pm$0.16, 0.76$\pm$0.08, 0.80$\pm$0.07, 0.74$\pm$0.09.

The points on the diagrams may be represented fairly well by the two straight lines *CB* and *DB* although the number of points available is not large enough to make certain that such a relationship is more than accidental. The fact, however, that a given change in the value of the horizontal component of the earth's magnetic field may accompany at one time a very small change in cosmic-ray intensity values and at another a very large change is not new in cosmic-ray work but is a well-known characteristic of magnetic storms. The variation of cosmic-ray intensity, as measured on the ground, may be widely different for two storms which produce similar magnetic effects.

In diagrams 3(*A*), (*B*), (*C*), (*D*) the line *AB*

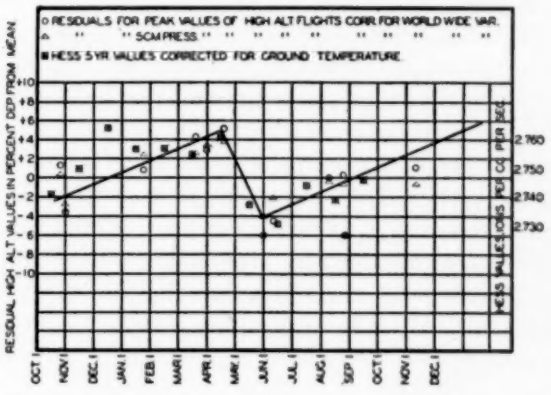


FIG. 4. Residuals from peak values and from values for 5 cm pressure, after correction for "world-wide" variations, plotted against time.

[6] V. F. Hess, Rev. Mod. Phys. **11**, 153 (1939).

has been drawn somewhat arbitrarily to represent the average trend of the plotted points. Such a line represents roughly the mean relation between Huancayo intensity values and high altitude values assuming a simple proportionality to exist. From the abscissa of each experimental point there was then subtracted the abscissa of the point on the line having the same ordinate value as the experimental point. Residual values thus obtained represent variations in the high altitude intensity values in excess of the mean variation to be expected as a result of changes in the Huancayo values. Thus these may reflect the effect upon the high altitude values of some additional disturbing factor. In Fig. 4 the the residuals thus derived from Figs. 3(*A*) and 3(*B*) are plotted as ordinates against time on the abscissa axis. A very similar curve may be obtained for the magnetic values by using in a like manner plots 3(*E*), (*F*), (*G*). As will be seen, the residual values increase continuously from October to April, drop suddenly to a minimum in the early summer and then increase slowly again. It seems difficult to fit the points on a sine curve, chiefly because of the very abrupt drop in the spring, so the two sets of points have been represented by a sort of jagged "saw-tooth" curve. Whether the variation shown here represents a true seasonal effect or not cannot be determined from the results of one year's experiments. The present experiments are being continued and should soon show whether the effect is a truly periodic one or not.

An interesting comparison may be made between the present results and those recently published by Hess.[7,8] The mean values from five years cosmic-ray intensity data taken on the Hafelekar (altitude 2300 m; 48° N mag. lat.) are shown as crossed circles in Fig. 4. Here the recorded cosmic-ray intensity values in ions per cc per sec. per atmosphere have been corrected by Hess for ground temperature variations and the resulting values are plotted against time. The values obtained by Hess follow very closely the "saw-tooth" curve representing the variation of the high altitude values. The percent variation is about eight times greater in the high altitude values. It should be noted that in the Hess data there has been no correction for "world-wide" variations as in the present results. However, presumably in a five-year mean the effect of such irregular variations would be less than in data taken over a shorter period of time.

The cause of the variations in the high altitude data as shown in Fig. 4, whether seasonal or nonseasonal, is not at once apparent. Theoretical calculations indicating a seasonal change at high altitudes have been made by Vallarta and Godart[9] and by others but the present variations do not agree well with such calculations.

It should be pointed out also that the curve in Fig. 4 does not seem to be a mere extension to high altitudes, with increased amplitude, of the seasonal effect obtained by many observers[4,6,10] at the surface of the earth. This latter curve shows a marked dependence upon air temperature with a maximum in January and a minimum in July. The phase therefore is in disagreement with the curve of Fig. 4. Thus the present experiments, within their limits of error, furnish no evidence that the seasonal effect observed in ground measurements results from any factor acting upon the cosmic rays outside of the earth's atmosphere.

If one wishes to enter the realm of speculation, one might attempt to explain the results shown in Fig. 3 in terms of variations in the earth's magnetic field combined with variations in the sun's magnetic field. Adopting such a viewpoint one might consider the variations of high altitude values along the line *DB* as caused by variations in the earth's magnetic field alone while variations along the line *CB* could be considered the combined result of simultaneous changes in the field of the sun and in the field of the earth. Epstein,[11] assuming a value of 25 gauss for the strength of the field at the magnetic poles of the sun, has shown as a result of such a field that electron rays below 1.0×10^9 ev would be completely eliminated in their passage to the earth, that rays in the range between 1.0×10^9 and 5.9×10^9 would be weakened in intensity and that rays of higher energies would be quite unaf-

[7] I am indebted to Professor Hess for making these results available to me in advance of their publication.
[8] V. F. Hess, Phys. Rev. **57**, 781 (1940).
[9] M. S. Vallarta and O. Godart, Rev. Mod. Phys. **11**, 180 (1939).
[10] A. H. Compton and R. N. Turner, Phys. Rev. **52**, 709 (1937).
[11] P. S. Epstein, Phys. Rev. **53**, 862 (1938).

fected. At the latitude of Chicago ($\lambda = 52°$N) where the threshold due to the earth's field is about 2.9×10^9 ev one might expect to find the cosmic-ray intensity at high altitudes subject to variation in either field. At high latitudes the threshold imposed by the sun's field would be the only one effective while at low latitudes the threshold imposed by the earth's field would be the predominant one.

Whether the sun's field varies in intensity as assumed above or even whether there is real evidence that the sun's field exists as a factor affecting cosmic rays are important questions still to be settled. Time variations of more than four percent in high altitude cosmic-ray intensities at Bismark, North Dakota ($\lambda = 56°$ N) reported by Millikan and Neher[3] suggest the possibility that such variations persist even up to very high latitudes where the best available data indicate a cosmic ray intensity independent of latitude. A similar variation of 2 percent[12] has just been reported in sea level values at magnetic latitude 58° N during a recent magnetic storm. Such changes would seem difficult to explain on the basis of a variation in the earth's field alone. These considerations, however, only emphasize the need of further systematic high altitude experiments at high latitudes.

Conclusions

(1) During the course of a year's high altitude measurements, changes as large as fourteen percent have been observed for the maximum ordinate of the ionization-pressure curve. The corresponding maximum change in the integrated area under the curves is about ten percent.

(2) The high altitude variations follow in general the "world-wide" variations already observed at ground stations which are believed to be caused by fluctuations in the magnetic field surrounding the earth.

(3) After correction for "world-wide" variations certain residual variations remain in the high altitude values which seem to be of significance. There is an indication of combined effects of the magnetic fields of the earth and sun although no fully satisfactory explanation can be at present advanced.

(4) The establishment of the existence of time variations such as are reported here is of importance for future high altitude measurements. Thus in determining the effect of latitude or longitude on high altitude cosmic-ray intensities it is very important to be sure that no time variation has taken place during the time elapsed between the flights at the different stations. Otherwise the result obtained may have but little relation to the effect which it is desired to measure.

The writer wishes to express his thanks to Professor A. H. Compton, who first suggested the present type of experiment, for his continued interest, to Dr. Elmer Dershem whose skill in the making of the electrometers used has contributed much to the success of the work and to Dr. John A. Fleming of the Carnegie Institution who furnished both the cosmic-ray and magnetic data. He also wishes to express his appreciation for the help of Messrs. Alois Bragagnolo, Leo Seren, Ted Novey and Warren Nyer who aided in the flights. This work has been supported in part by a grant from the Penrose Fund of the American Philosophical Society and in part by the generosity of Mrs. Henry A. Strong.

[12] D. H. Loughridge and P. E. Gast, Phys. Rev. **57**, 938 (1940).

Fine Structure in the Directional Intensity of Cosmic Rays

DENSIL COOPER
University of Missouri, Columbia, Missouri

(Received June 6, 1940)

The effect relating to fine structure in the zenith-angle distribution of cosmic-ray intensity was investigated with a triple coincidence circuit of Geiger-Müller counters. Three counters, each of diameter 8.8 cm, active length 36 cm, were made according to the method of Shonka. The counters were used with a separation of 50 cm between the centers of adjacent counters in a modified circuit of the Neher-Harper type. Readings were taken in a cyclic fashion to minimize errors due to possible instrumental sensitivity drifts, changes in the barometric pressure, and changes in the magnetic field of the earth. Five-degree intervals were investigated in the zenith-angle range from 0° to 45°, inclusive, in all the directions studied. The survey of intensities in the east, west, north, south, southwest, southeast, northwest, and northeast azimuths exhibits intensity patterns with small oscillations. The prominences measured in terms of the largest positive deviations from the $\cos^2\theta$-curve are of the order of two or three percent of the intensity, and they occur at approximately 7°, 20°, and 37°. The directional intensity pattern shows an approximate symmetry about the zenith. These results tend to confirm the predictions of Schremp relative to the existence of a fine structure due to a line or banded nature for the energy spectrum for cosmic rays at infinity.

INTRODUCTION

SCHREMP[1,2] developed the theory of fine structure in the directional intensity of cosmic rays, i.e., prominences and depressions in the intensity curve plotted as a function of the zenith angle. He inferred that a fine structure might arise from: (1) magnetic effects associated with the edges of the "main" cones and the "penumbral" bands of the Lemaître-Vallarta[3,4] theory, which would give rise to patterns of asymmetrical loci of intensity prominences about the zenith; or, (2) absorption effects associated with the traversal of range limits of lines or bands in the energy spectrum for cosmic rays at infinity, which would provide a symmetrical pattern. The data from Johnson's[5] world-wide asymmetry measurements supplied evidence that a fine structure was present, despite the fact that the experimental conditions and methods would tend to mask the fine structure. In 1939, Ribner[6] and Cooper,[7] independently, reported results on fine-structure measurements that demonstrated the existence of the effect. Ribner's observations in the east and west azimuths showed positional symmetry of the fine structure. This experiment, which has been performed at Columbia, Missouri, supplies further information in the east and west azimuths and extends the observations to other azimuths.

APPARATUS

Three Geiger-Müller counter tubes were arranged for triple coincidence counting. Each counter had a copper cylindrical cathode 8.8 cm in diameter, and a tungsten central wire 0.005 in. in diameter. These parts were sealed in glass, and the counter was cleaned, baked, and filled with hydrogen according to the methods of Shonka.[8] The characteristics of the counters are shown in Fig. 1, where the counting rate for each counter is plotted as a function of the applied voltage. The counters were operated at 1520 volts, which is well within the plateau range for each counter. The counters exhibited "clean" pulses, and tests with gamma-rays showed them to be highly efficient.

A diagram of the circuit is shown in Fig. 2. The first stage is of the Neher-Harper[9] type. The second stage lengthens and changes the sign of the pulse. Coincidences operate the thyratron.

[1] E. J. Schremp, Phys. Rev. **53**, 915A (1938); **54**, 157 (1938).
[2] E. J. Schremp and H. S. Ribner, Rev. Mod. Phys. **11**, 149 (1939).
[3] G. Lemaître and M. S. Vallarta, Phys. Rev. **49**, 719 (1936).
[4] R. A. Hutner, Phys. Rev. **55**, 15 (1939); **55**, 614 (1939).
[5] T. H. Johnson, Phys. Rev. **45**, 569 (1934); **48**, 287 (1935).
[6] H. S. Ribner, Phys. Rev. **55**, 127 (1939); **56**, 1069 (1939).
[7] D. Cooper, Phys. Rev. **55**, 1272 (1939).
[8] J. B. Hoag, *Electron and Nuclear Physics* (D. Van Nostrand Company, New York, 1938), p. 431.
[9] H. V. Neher and W. W. Harper, Phys. Rev. **49**, 940 (1936).

The final stage including the strobotron tube[10] was added as a Western Electric 5M message register will operate in the plate circuit of this tube. Small Edison type syd batteries supplied the high voltage, which was kept constant within 4 volts. The 250-volt plate voltages were supplied by two RCA type TMV-118-B power supply units. The various grid biases were obtained from battery units.

The counter train was mounted in a vertical plane on a large semicircular track, which was graduated in five-degree intervals on the track. The angular adjustments could be made to approximately three minutes of arc. The counters were separated 50 cm between adjacent counters. Under these conditions, the maximum angle of divergence of the cosmic-ray beam in the plane of rotation of the counter train was 10°, and the corresponding angle perpendicular to the plane of rotation of the counter train was 40°. However, the actual angular resolution is much better than the dimensions indicate. It can be shown from the geometry of the set-up that for a maximum angle of divergence of 10° in the plane of rotation of the counter train, more than one-half of the rays are incident through the angle within 1.4° on each side of the zenith angle. More than one-half of the rays are incident through an angle of 5.8° on each side of the plane of rotation. The counting rate was approximately 6 counts per minute with this separation of the counters. Eckart and Shonka's[11] formula gave a value of 0.00006 "accidental" coincidence per second.

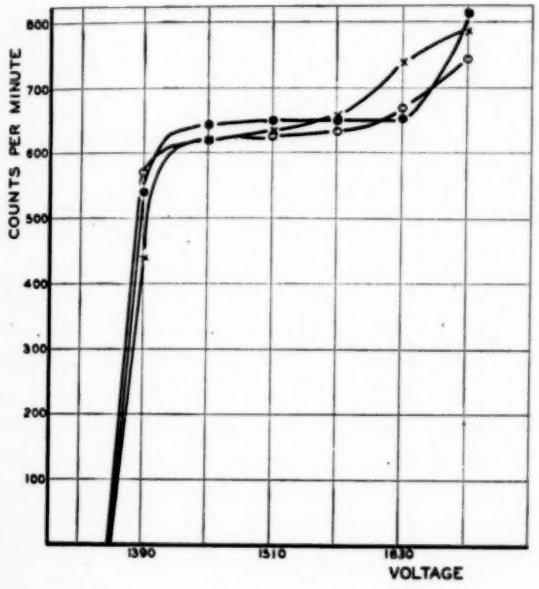
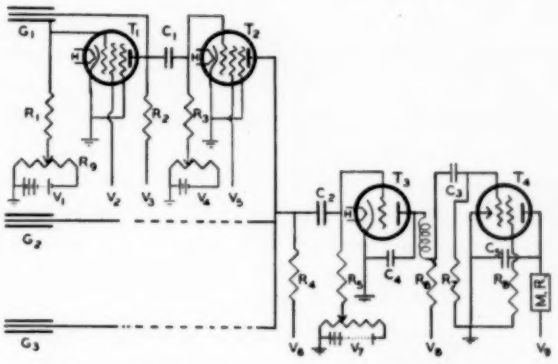


FIG. 1. Characteristic curves for the Geiger-Müller counters with the counting rate plotted as a function of the anode voltage.

[10] N. S. Gingrich, Rev. Sci. Inst. **7**, 207 (1936).

OBSERVATIONS

The observations were made in the attic of the Physics Building at an altitude of 760 ft. above sea level and a geomagnetic latitude[12] of 49° N. Only a thin, uniform slate roof shielded the apparatus from the sky. Readings were taken at every five degrees of zenith angle from 0° to 45°, inclusive, in each azimuth that was investigated. It was necessary to use the "method of cycles" in collecting the data to minimize the effect of: (1) changes in the barometric conditions, especially changes in barometric pressure; (2) changes in the intensity of the magnetic field of the earth; and, possibly, (3) slight sensitivity drifts in the triple coincidence apparatus. The changes in the intensity arising from changes in the barometric pressure and changes in the magnetic field of the earth are not small



$G_1 = G_2 = G_3$—Geiger counters
$T_1 = T_2$—Type 57
T_3 = 885 Thyratron
T_4 = Strobotron
V_1 = 10 volts
V_2 = 45 volts
V_3 = 1520 volts
V_4 = 2 volts
V_5 = 90 volts
V_7 = −90 volts
$V_8 = V_3 = V_9$ = 250 volts
H = 2.5 volts—a.c.

$R_1 = 10^7$ ohms
$R_2 = 2 \times 10^6$ ohms
$R_3 = 5 \times 10^4$ ohms
$R_4 = R_7 = R_8 = 10^5$ ohms
$R_5 = 5 \times 10^4$ ohms
$R_6 = 10^5$ ohms
$R_9 = 2 \times 10^6$ ohms
$C_1 = 50 \mu\mu f$
$C_2 = 0.02 \mu f$
$C_3 = 0.1 \mu f$
$C_4 = 0.5 \mu f$
$C_5 = 6 \mu f$

FIG. 2. Modified triple coincidence circuit with the Neher-Harper first stage.

[11] C. Eckart and F. R. Shonka, Phys. Rev. **53**, 752 (1938).

[12] A. H. Compton, Phys. Rev. **43**, 387 (1933).

TABLE I. *The analysis of the data. Cycle No. 48—North (15 min. of observation at each angle).*

Zenith Angle (θ)	0°	5°	10°	15°	20°	25°	30°	35°	40°	45°
$N_{48}(\theta)$	6.20	6.40	5.93	6.07	6.20	4.33	4.80	3.93	2.60	2.80
$N_{48}(\theta)\sec^2\theta$	6.2	6.5	6.1	6.5	7.0	5.3	6.4	5.8	4.4	5.6
									$N_{48}(0)=6.0$	
$I_{48}(\theta)$	1.03	1.07	0.99	1.01	1.03	0.72	0.80	0.66	0.43	0.47

$N_{48}(\theta)$ is the counts per minute at the angle θ.
$N_{48}(0)$ is the average value of $N_{48}(\theta)\sec^2\theta$.
$I_{48}(\theta)$ is $N_{48}(\theta)/N_{48}(0)$.

$I(\theta)=\frac{1}{n}\sum_{i=1}^{n} I_i(\theta)$, where $I(\theta)$ is the normalized intensity in a particular azimuth, and n is the number of cycles in this azimuth.

$\Delta(\theta)=I(\theta)-\cos^2\theta$, where $\Delta(\theta)$ is the deviation from the $\cos^2\theta$ curve.

(For convenience in comparison, $\Delta(\theta)$ has been shifted to zero in each azimuth.)

factors; in percentage change of the intensity, they compare in magnitude[12,13] with the fine structure prominences. By completing in a relatively short time a cycle of the angles in the azimuth under study, these factors which are beyond experimental control remained essentially constant during the cycle. Theoretically, the error arising from these sources may be reduced to any desired value by increasing the number of cycles. By treating each cycle as a separate entity, only those sensitivity changes in the coincidence apparatus that occurred during the cycles affected the final result. The method of analyzing the data is illustrated in Table I. Since the intensity is approximately proportional to $\cos^2\theta$, where θ is the zenith angle, $N(\theta)\sec^2\theta$ has nearly a constant value throughout the cycle and the average value of this quantity for the cycle has been used to normalize the intensity curve for that cycle. It is more convenient to

TABLE II. *Directional intensity data. Zenith angle, θ; deviations from $\cos^2\theta$, $\Delta(\theta)$; probable error, ϵ.*

θ	East azimuth[1] $\Delta(\theta)$	ϵ	West azimuth[2] $\Delta(\theta)$	ϵ	North azimuth[3] $\Delta(\theta)$	ϵ	$\Delta(\theta)$ for 68 cycles	ϵ	South azimuth[4] $\Delta(\theta)$ 60 regular cycles	ϵ	$\Delta(\theta)$ 8 anomalous cycles	ϵ
0	0	0.008	0	0.010	0	0.009	0	0.009	0	0.009	0	0.025
5	+0.009	0.008	+0.020	0.011	+0.031	0.008	+0.015	0.008	+0.019	0.008	−0.017	0.024
10	+0.002	0.008	+0.034	0.011	+0.005	0.009	+0.007	0.009	+0.012	0.009	−0.034	0.024
15	−0.046	0.007	−0.008	0.009	−0.021	0.009	−0.007	0.008	−0.016	0.009	+0.056	0.024
20	−0.003	0.008	+0.006	0.012	−0.007	0.007	−0.003	0.009	+0.008	0.009	−0.088	0.024
25	−0.053	0.008	−0.006	0.010	−0.061	0.009	−0.015	0.008	−0.014	0.009	−0.016	0.023
30	−0.038	0.007	−0.036	0.009	−0.034	0.007	−0.015	0.008	−0.026	0.009	+0.017	0.022
35	+0.019	0.009	−0.025	0.014	−0.038	0.008	−0.014	0.007	−0.012	0.008	−0.038	0.021
40	−0.032	0.008	−0.016	0.010	−0.060	0.006	−0.028	0.006	−0.024	0.006	−0.008	0.020
45	−0.025	0.008	−0.023	0.011	−0.057	0.006	−0.029	0.006	−0.031	0.006	−0.011	0.018

θ	Southwest azimuth[5] $\Delta(\theta)$	ϵ	Southeast azimuth[6] $\Delta(\theta)$	ϵ	Northwest azimuth[7] $\Delta(\theta)$	ϵ	Northeast azimuth[8] $\Delta(\theta)$	ϵ
0	0	0.014	0	0.011	0	0.011	0	0.009
5	+0.041	0.012	+0.037	0.009	+0.001	0.011	+0.004	0.011
10	+0.040	0.015	+0.031	0.012	+0.034	0.011	+0.049	0.014
15	−0.037	0.012	−0.027	0.009	−0.002	0.010	−0.027	0.010
20	+0.015	0.011	−0.003	0.009	+0.009	0.011	−0.028	0.009
25	−0.034	0.012	−0.017	0.009	−0.044	0.010	+0.004	0.008
30	−0.023	0.013	−0.066	0.009	−0.025	0.009	−0.055	0.008
35	−0.022	0.011	−0.048	0.009	−0.036	0.009	−0.044	0.009
40	−0.021	0.011	−0.042	0.008	−0.039	0.011	−0.047	0.009
45	−0.033	0.011	−0.043	0.009	−0.049	0.007	−0.070	0.009

[1] 40 cycles from 0° to 45°, plus 20 cycles from 0° to 30°; 15-min. observations at each angle during each cycle.
[2] 20 cycles from 0° to 45°, plus 20 cycles from 0° to 30°; 15-min. observations at each angle during each cycle.
[3] 60 cycles from 0° to 45°; 15-min. observations.
[4] 68 cycles from 0° to 45°; 15-min. observations; 8 cycles run in 2 consecutive days in August showed a shift of the fine structure pattern. These 8 cycles are listed separately.
[5] 30 cycles from 0° to 45°; 10-min. observations.
[6] 30 cycles from 0° to 45° with 10-min. observations, plus 20 cycles from 0° to 45° with 15-min. observations; results weighted accordingly.
[7] 40 cycles from 0° to 45°; 15-min. observations.
[8] 40 cycles from 0° to 45°; 15-min. observations.

[12] S. E. Forbush, Phys. Rev. **54**, 975 (1938).

plot the deviations $\Delta(\theta)$ from the empirical $\cos^2$-curve than the zenith-angle curve. This representation flattens out the distribution curve, and shows any prominences and depressions as oscillations about a perfect $\cos^2$-distribution.

The results of the survey of the directional intensities in eight azimuths are tabulated in Table II, and $\Delta(\theta)$ is plotted against θ for the various azimuths in Fig. 3. The azimuths were investigated in the order in which they are listed in the table. The observations were made during the time from May 15, 1939, to January 15, 1940.

The probable errors were computed by the "method of residuals" from the value of $I(\theta)$ by the relation

$$\epsilon=0.67\left[\sum_{i=1}^{n}\frac{r_i^2}{n(n-1)}\right]^{\frac{1}{2}},$$

where n is the number of cycles, r_i is the ith residual, and ϵ is the probable error of the average value. The probable errors for the results of eight anomalous cycles in the south (shown in Fig. 3) were computed from the total counts by the formula

$$\epsilon=0.67\sqrt{C}/N(0)T,$$

where C is the total number of counts, and T is the time of observations.

Discussion of Results

The results of Ribner's[6] independent survey of intensities in the east and west azimuths are shown in Fig. 4. A comparison of the two surveys shows qualitative agreement and reasonably good quantitative agreement.

In the present work, 68 cycles were taken in the south. Sixty of these cycles exhibited a fine structure quite comparable in magnitude and position with the fine structure in the north (as is shown in Fig. 3). However, eight anomalous cycles, taken during two consecutive days in August, gave a complete inversion of the fine-structure pattern. Short period shifts of this type were observed by Ribner; and Schremp[14] has recently reported studies of the time variation of the directional intensity pattern. These shifts have been attributed to changes in the barometric or the magnetic conditions. It is interesting to note that the eight anomalous cycles were made during a period of activity of the aurora borealis. Similar shifts of shorter duration are believed to be present in the data. Because of the time shifts there is some question whether definite magnitudes can be ascribed to the prominences, even with a large number of counts.

The directional intensity curves of Fig. 3 show approximate symmetry of the fine structure

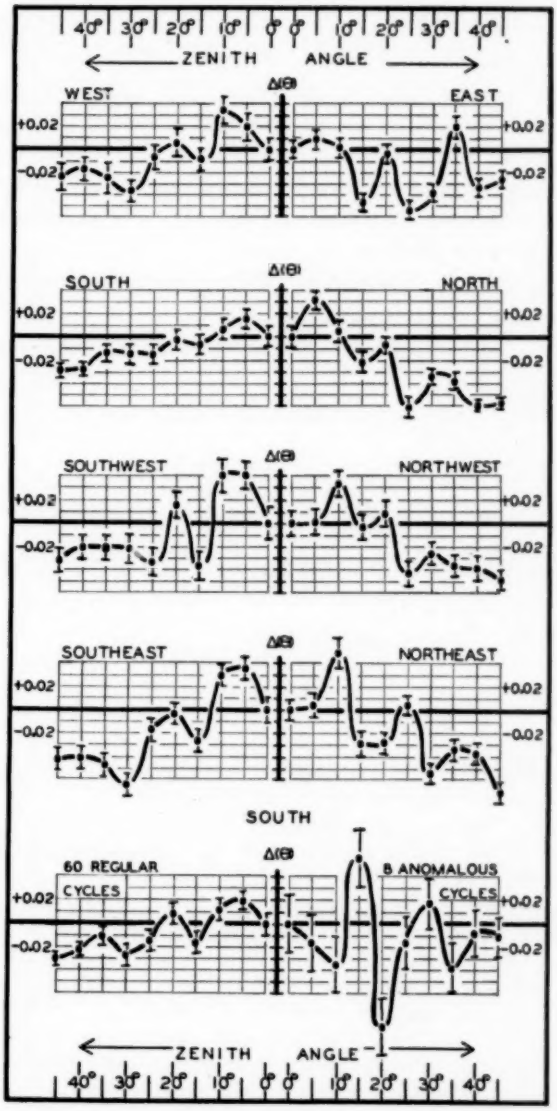


FIG. 3. Directional intensity patterns for eight azimuths. The deviations $\Delta(\theta)$ of the normalized directional intensity from the empirical $\cos^2$-curve are plotted as a function of the zenith angle θ, in degrees. The first eight graphs show the approximate positional symmetry of the fine structure in the various azimuths. The observed shift of the directional intensity pattern is shown in the last two graphs.

[14] E. J. Schremp, Phys. Rev. **57**, 1061A (1940).

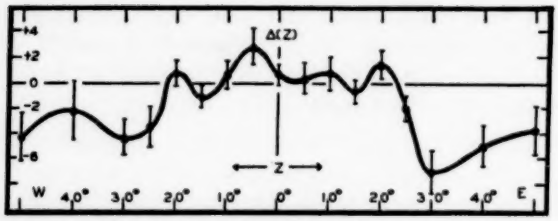


FIG. 4. Ribner's directional intensity results in the east and west azimuths. z is the zenith angle in degrees. $\Delta(z)$ is the absolute deviation, in percent of the zenith angle value, of the observed intensity from that given by the $\cos^2 z$ relation.

about the zenith with intensity prominences in the zenith-angle range from 5° to 10°, near 20°, and in the range from 35° to 40°. These results tend to confirm the predictions of Schremp relative to the existence of a fine structure due to a line or banded nature of the energy spectrum for cosmic rays at infinity. That the fine structure may not arise partly from magnetic effects is not entirely precluded, as the penumbral bands of the Lemaître-Vallarta theory form a sort of "C" and might give rise to a perturbation of the fine structure due to absorption without entirely destroying the approximate symmetry of the pattern. But as Columbia, Missouri is near the "knee" of the geomagnetic latitude intensity curve and the pattern shows an approximate symmetry about the zenith, the most plausible explanation in terms of primary cosmic rays seems to be that there are lines or bands in the energy spectrum at infinity. Then, it is necessary that these lines or bands be successively absorbed out by range absorption by increasing the atmospheric path traversed by the primary radiation. This explanation of the fine structure in terms of absorptive effects is not in discord with the present results, Ribner's results, and the deductions from Johnson's data. Absorption experiments are in progress to test this interpretation of the experimental results.

The author wishes to express his deepest appreciation to Professor N. S. Gingrich whose keen advice, earnest cooperation, and inspiring stimulation were invaluable throughout the course of this investigation. He also wishes to express his gratitude to Professor E. J. Schremp for many invaluable suggestions and constructive criticisms.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

Multiple Scattering of Fast Electrons and Alpha-Particles, and "Curvature" of Cloud Tracks Due to Scattering

E. J. WILLIAMS
University College of Wales, Aberystwyth, Great Britain
(Received October 23, 1939)

According to a number of observers the multiple scattering of fast electrons by thin foils is appreciably less than the scattering predicted by the theory of this effect given some time ago by the present author. The statistical part of the theory, which in the earlier work was developed with special regard to the scattering of cosmic-ray particles, has been reconsidered more closely for the conditions of the experiments on fast electrons. The results confirm within a few percent the scattering given by the earlier general formula. The discrepancies must accordingly be attributed to experimental error or to a failure of the basic collision theory. A new formula is given which represents the mean projected deflection within 1 percent over the whole range of experimental conditions. The *distribution* of the projected deflections is considered more closely than in the earlier paper, and a general expression for the most probable deflection in space is also given. The theory is extended to the multiple scattering of α-particles, the quantum-mechanical collision theory in this case taking the form of classical mechanics. The results are in satisfactory agreement with the early experiments of Geiger and of Mayer. A discussion is given of recent papers on the subject. Reasons are given for the nonoperation of the interference effect considered by Wheeler, and which according to him appreciably reduces the scattering. The numerical results of a new treatment of multiple scattering by Goudsmit and Saunderson are shown to be the same as those required by the general formula given by the writer's theory. Other points raised by these authors are also discussed. A summary of the theoretical results is given in Section (7).

IN a recent publication[1] (1939) the writer discussed the problem of the multiple scattering of cosmic-ray particles and fast electrons. The theoretical value of the mean deflection of cosmic-ray particles by metal plates was shown

[1] E. J. Williams, Proc. Roy. Soc. **169**, 531 (1939).

to be in very good agreement with the experimental results obtained by Blackett and Wilson[2] (1938). However, results obtained by Fowler and Oppenheimer[3] (1938) indicated that the scattering of electrons with energy of the order 10 Mev by thin foils (~0.01 cm lead) is appreciably less than the calculated value. In view of this discrepancy, and of further similar experiments in progress, I have considered more closely the multiple scattering under the conditions concerned, *viz.* thin foils as scatterers rather than thick plates as in the cosmic-ray experiments—the earlier calculations being primarily concerned with the latter. The results of these more specific calculations for thin foils are practically identical with those required by the general formula given in the earlier paper (which we shall refer to as S). The discrepancy between the above experiments on fast electrons and theory thus remains unchanged. The multiple scattering of α-particles, which was not considered in the earlier paper, is in satisfactory agreement with the theory. In this connection it might be mentioned that while the statistical part of the theory is exactly the same for α-particles as for fast electrons and cosmic-ray particles, the basic collision theory is different in that classical mechanics must be used and not the Born approximation. In this paper we give: (1) an expression for the mean deflection which while numerically identical with the earlier formula for the scattering of cosmic-ray particles is slightly more accurate for very large departures from such conditions, (2) results of more detailed calculations on the *distribution* of scattering. (In the earlier paper the distribution was not calculated with the same accuracy as the mean deflection. The new results, however, show that the division of the distribution into a Gaussian distribution and a single-scattering "tail" as given earlier is remarkably accurate.) (3) a fuller discussion of the effect of inelastic electronic scattering, (4) a fuller discussion of the effect of the "geometrical" limitation on the tracks observed in the experiments, (5) a discussion of the effect of the scattering of the tracks in the gas of the cloud chamber, and the "curvature" of cloud tracks produced by scattering, (6) an extension of the theory to α-particles and comparison with experiment, (7) a summary of results, (8) a discussion of some recent papers on the subject. (6), (7) and (8) were added in proof (April 12, 1940).

[2] P. M. S. Blackett and J. G. Wilson, Proc. Roy. Soc. **160**, 304 (1937).

[3] W. A. Fowler and J. Oppenheimer, Phys. Rev. **54**, 320 (1938).

(1) Mean Deflection

As in the earlier paper we shall, throughout this one, use as a unit of angle, the quantity

$$\delta = 2Zze^2(Nt)^{\frac{1}{2}}/M_p\beta^2c^2\xi = 2Ze(Nt)^{\frac{1}{2}}/\beta H\rho. \quad (1)$$

N is the number of atoms per cc of scattering foil, Z the atomic number, t the thickness of the foil, M_p the mass of the scattered particle, ze its charge, βc its velocity, $\xi=(1-\beta^2)^{-\frac{1}{2}}$, and ρ is the curvature of the path of the particle in a magnetic field H. In terms of δ as unit the mean projected deflection given by Eq. (37) of the earlier paper is

$$\bar{\alpha} = 3.69 + 0.28 \log_{10} Z^{4/3}z^2Dt/W\beta^2, \quad (2)[4]$$

where D is the density and W the atomic weight of the scattering material. This result is based on a detailed calculation of $\bar{\alpha}$ for a special case, *viz.* the scattering of fast particles ($\beta=1$) with unit charge ($z=1$) by 1 cm of lead, and a separate calculation of the rate of change of $\bar{\alpha}$ with the conditions. The results of the analysis may, however, be expressed in a different form, which avoids a certain second-order approximation which was made in deducing (2). This alternative expression is

$$\bar{\alpha} = 0.80\bar{\alpha}_1 + 1.45 \quad (3)$$

$$\bar{\alpha}_1^2 = \log_e (2\pi NtZ^{4/3}z^2\hbar^2/3\cdot 1m^2\beta^2c^2), \quad (3a)$$
$$= \log_e M \text{ say.}$$
$$= 7.45 + 2.3 \log_{10} (Z^{4/3}z^2Dt/W\beta^2).$$

(2) and (3) are identical for the scattering of fast particles by 1 cm of lead ($M_0 = 3.3\times10^4$),[5] they differ by only 2 percent for $M=300$ (corresponding to the conditions in the experiments on the scattering of electrons by thin foils) and they differ by 8 percent for $M=30$. That Eq. (3) is

[4] This assumes that the de Broglie wave-length of the scattered particle is much larger than the dimensions of the scattering nuclei. This is fully satisfied for the experiments under consideration, i.e. scattering of electrons with energy of the order of 10 Mev or less. The scattering when this condition is not satisfied was fully dealt with in the earlier paper.

[5] For $M=M_0$, $\bar{\alpha}_1^2=10.44$ so that in general $\bar{\alpha}_1^2 = 10.44 + 2.3 \log_{10} M/M_0$, i.e., $\bar{\alpha}_1 = 3.23\{1 + (2.3/10.44) \log_{10} (M/M_0)\}^{\frac{1}{2}}$. The expression (2) follows if we put the square root of the factor within brackets equal to $1+\frac{1}{2}(2.3/10.44) \log_{10} (M/M_0)$, and then substitute in (3).

more accurate than (2) in the region where they differ ($M \ll M_0$) was found by specific calculations of $\bar{\alpha}$ for $M \ll M_0$, *viz.* $M=5$, 15, 33, 330, and 3300, respectively, similar to the earlier calculation for $M=M_0$, $=3.3\times10^4$.* These calculations show that provided M is greater than about 30 ($\sim$0.001 cm lead, $\beta=1$) Eq. (3) is correct within 1 or 2 percent, and it is only when M is less than about 10 that the equation starts to become appreciably wrong. For such small M the shielding of the nuclear field affects the scattering through even the largest angles which contribute to the multiple scattering, and the error in (3) for $M<10$ is in fact more due to this than to the change in the statistical conditions. Regarding the latter it might be pointed out that M actually measures the average number of scattering collisions experienced by the electron in traversing the foil, and the condition $M \gg 1$ is essential to ensure multiple scattering (Eq. (32a) of S). If $M \ll 1$ then the scattering is single for all angles however small,—the probability of experiencing two deflections in traversing the foil being of the order of M^2. Excepting those on α-particles, no experiments have been carried out with foils of thickness corresponding to M less than about 10.

(2) Distribution of Scattering

In the earlier paper the distribution of the deflections produced by the scatterer was represented by

$$P(\alpha)=G(\alpha)+S(\alpha), \tag{4}$$

where $G(\alpha)$ is a Gaussian distribution with mean deflection α_m, and $S(\alpha)$ is a single scattering distribution starting at ϕ_2 and equal to π/α^3. Here ϕ_2 is the angle at which a Gaussian distribution with mean angle $=\bar{\alpha}_1$ (Eq. (3a)) intersects the single scattering curve π/α^3. This intersection is given with sufficient accuracy by the empirical relation

$$\phi_2=5.1\bar{\alpha}_1-4.0. \tag{5}$$

α_m is chosen so that the composite curve gives the correct mean deflection $\bar{\alpha}$ (Eq. (3)),—which it must be noted has been calculated without making the approximation (4) regarding the distribution. The mean deflection due to the distribution π/α^3 beyond ϕ_2 is

$$\int_{\phi_2}(\pi/\alpha^3)\alpha d\alpha=\pi/\phi_2,$$

while the corresponding area is

$$\int_{\phi_2}(\pi/\alpha^3)d\alpha=\pi/2\phi_2{}^2.$$

Table I. *Values of $G(\alpha)$ and $S(\alpha)$.*

α	$G(\alpha)$	$S(\alpha)$	$P(\alpha)$ $=G+S$	$P(\alpha)$ (new calc)
0	0.201	0	0.201	0.208*
2	0.176	0	0.176	0.179
4	0.118	0	0.118	0.115
6	0.061	0	0.061	0.058
8	0.0239	0	0.0239	0.0248
$8.3=\phi_2$	0.0204	0; 0.0055	0.0231	0.0218
10	0.0072	0.0031	0.0103	0.0098
12	0.0017	0.0018	0.0035	0.0042
14	0.0003	0.0012	0.0015	0.0020
$>2\phi_2$	negligible	π/α^3	π/α^3	$(\pi/\alpha^3)(1+3\pi\alpha_m{}^2/\alpha^2)$

* The value at $\alpha=0$ is given by the expression:
$(2/\pi\alpha_1')[0.75+0.25(e^{-2/\alpha_1'}-(2/\alpha_1')\int_{2/\alpha_1'}^{\infty}e^{-y}y^{-1}dy)]$.

It follows that in order to give a resultant curve of unit area and with the correct mean deflection, $\bar{\alpha}$,

$$\alpha_m=(\bar{\alpha}-\pi/\phi_2)\div(1-\pi/2\phi_2{}^2) \tag{6a}$$

and

$$G(\alpha)=(1-\pi/2\phi_2{}^2)(2/\pi\alpha_m)\exp(-\alpha^2/\pi\alpha_m{}^2). \tag{6b}$$[6]

* These are described in Appendix B of the earlier paper (S). The calculations involve very little numerical computation. For large M expressions for upper and lower limits to $\bar{\alpha}$ which differ by only a few percent may in fact readily be obtained in terms of standard tabulated integrals. With the notation of Section (2) of this paper the mean deflection due to collisions giving projected deflections $\phi<\gamma=2\phi_1$ is $\alpha_1'=(\log 4M)^{\frac{1}{2}}$, and the distribution is practically Gaussian. Only in a fraction of the traversals of the scattering foil equal to $1-e^{-\frac{1}{2}}$, $=0.221$, do collisions giving deflections greater than $2\phi_1$ take place. The effect of $\phi>2\phi_1$ in these traversals is obtained by the method described in the above-mentioned appendix. We find by that method that the resultant deflection $\bar{\alpha}$ has the limits

$$\bar{\alpha}\text{ (lower limit)}=0.779\alpha_1'+0.221\alpha_2'$$
$$\text{(upper limit)}=0.75\alpha_1'+0.25\alpha_2'$$

where

$$\alpha_2'=\int_{\sqrt{2}/\alpha_1'}^{\infty}e^{-x^2}\{(x\alpha_1')^{-1}+\tfrac{1}{2}x\alpha_1'-1\}dx.$$

The numerical value of this integral for a given α_1', i.e., given M, may be obtained from standard tables. The limits differ by 2 percent for $M=3\cdot10^4$, and only by 4 percent even for M as low as 15. The calculations in Appendix B of S, which were based on a division of the collisions at $\gamma=\phi_1$, give limits rather further apart. Mainly to check the arithmetic the new calculations of $\bar{\alpha}$ for the different M quoted above were carried out on the basis of $\gamma=2\phi_1$ as well as $\gamma=\phi_1$. The results agree within 1 percent.

[6] The normalizing factor $(1-\pi/2\phi_2{}^2)$ was omitted in the earlier paper, being very nearly unity under the conditions of the cosmic-ray experiments (actually $=0.992$).

The theoretical distribution is tabulated below for $M=330$, which is typical of the experiments on the scattering of fast electrons by thin foils (corresponding to 0.01 cm of lead). For this case $\bar{\alpha}_1=(\log_e M)^{\frac{1}{2}}=2.41$. $\phi_2=5.1\bar{\alpha}_1-4.0=8.3$. $\bar{\alpha}=0.80\bar{\alpha}_1+1.45=3.40$. $\alpha_m=(\bar{\alpha}-\pi/\phi_2)/(1-\pi/2\phi_2^2)=3.09$. The second column of Table I gives the values of $G(\alpha)$, and the third the values of S, which is zero up to ϕ_2, and equal to π/α^3 beyond ϕ_2. The fourth column is the sum of the second and third and represents the resultant distribution. The last column represents the results of a more accurate and detailed calculation of the distribution, as described below. It will be seen that the two distributions lie very close together, and in particular that the Gaussian distribution with mean angle given by (6b) is adequate up to ϕ_2.

The distribution represented in the last column was obtained as follows. Let $Q(\alpha)$ represent the distribution of α due to single scattering through $\phi<\gamma$, and let $R(\alpha)$ represent the distribution due to $\phi>\gamma$. Q and R refer to the algebraic value of α and not its absolute value. Since the resultant deflection is the algebraic sum of the deflection due to $\phi<\gamma$, and the deflection due to $\phi>\gamma$, the resultant distribution is

$$P(\alpha)=\int_{-\infty}^{+\infty}Q(x)R(\alpha-x)dx. \tag{7}$$

In the earlier paper (in calculating the mean deflection) a division of ϕ at $\gamma=\phi_1$ (Eq. (28) of S) was made, and it was shown (page 568 of S) that the distribution $Q(\alpha)$ due to $\phi<\phi_1$ is Gaussian within a few parts in 10,000, having a mean deflection $\bar{\alpha}_1=(\log_e M)^{\frac{1}{2}}$ (Eq. (3a)). The angle ϕ_1 is so defined that in traversing the foil the electron suffers, on the average, one collision with $\phi>\phi_1$, i.e.

$$\int_{\phi_1}(\pi/\phi^3)d\phi=1 \tag{8}$$

giving $\phi_1=(\pi/2)^{\frac{1}{2}}$. If we take $\gamma=2\phi_1$ the departures of $Q(\alpha)$ from Gaussian, though about four times greater, are still less than 1 percent and may be neglected. The mean deflection for this Gaussian is $\bar{\alpha}_1'=(\log_e 4M)^{\frac{1}{2}}$, so that $Q(\alpha)=(\pi\bar{\alpha}_1')^{-1}\exp\left[-\alpha^2/\pi\bar{\alpha}_1'^2\right]$. The average number of collisions with $\phi>2\phi_1$ is

$$\int_{2\phi_1}(\pi/\phi^3)d\phi=(\pi/8\phi_1^2)=\tfrac{1}{4}.$$

Hence for $e^{-\frac{1}{4}}$, $=0.779$, traversals of the foil no collisions with $\phi>2\phi_1$ occur, so that $R(\alpha)$ is a delta-function at $\alpha=0$, and $P(\alpha)=Q(\alpha)$. For $\frac{1}{4}e^{-\frac{1}{4}}=0.195$ traversals, one collision with $\phi>2\phi_1$ occurs, and $R(\alpha)=(2\pi/|\alpha|^3)$ for $\alpha>2\phi_1$ and $\alpha<-2\phi_1$, $R(\alpha)=0$ for other α. In a fraction $(\frac{1}{4})^2(\frac{1}{2})e^{-\frac{1}{4}}=0.024$ of traversals, two collisions through $\phi>2\phi_1$ occur. This fraction is, however, so small that no error greater than 1 percent can arise if we group these 0.024 traversal together with 0.024 traversal having no collision with $\phi>2\phi_1$, and treat them as 0.048 traversal having one collision with $\phi>2\phi_1$. Three or more collisions with $\phi>2\phi_1$, which occur only on 0.002 occasion, may be similarly treated. On this basis

$$P(\alpha)=0.75Q(\alpha)+0.25\int_{-\infty}^{+\infty}Q(x)R(\alpha-x)dx,$$

where $R=(2\pi/|\alpha|^3)$ for $\alpha<-2\phi_1$, and $>2\phi_1$, $R=0$ for $-2\phi_1<\alpha<2\phi_1$. The integral may be reduced to a power series in $(\alpha/\bar{\alpha}_1')$, which converges fairly rapidly for the values of α concerned.[7] For the case represented in Table I, i.e., $M=330$, we have $\bar{\alpha}_1'=(\log_e 4\times330)^{\frac{1}{2}}=2.68$, and $2\phi_1=(2\pi)^{\frac{1}{2}}=2.50$. $P(\alpha)$ has been evaluated for this case, and the results are given in the last column of the table. It might be added that the mean deflection corresponding to the last column is represented within 1 percent by the general formula (3).

(3) Effect of Inelastic Electronic Scattering

In the earlier paper it was pointed out that the intensity of the inelastic scattering by an atomic electron is for small angles given by the Rutherford formula for a nucleus of unit charge, with an effective lower limit at an angle of the order of $\lambda\!\!\!^{-}/a$ where $2\pi\lambda\!\!\!^{-}$ is the de Broglie wavelength of the scattered electron and a represents the dimensions of the orbit of the atomic electron. This cut-off at $\lambda\!\!\!^{-}/a$ is in the same region

[7] I am indebted to Professor Goudsmit for pointing this out to me, and supplying the formula concerned.

of angle as the cut-off of the elastic scattering by shielding. Since the multiple scattering is very insensitive to the value of the effective lower limit the inelastic scattering, which in any case has only a small effect, may be assumed to have the same cut-off as the elastic scattering. The inelastic scattering has a cut-off also at large angles, the maximum angle of scattering of an electron of energy $(\xi-1)mc^2$ by another electron being approximately $(2/\xi)^{\frac{1}{2}}$, $=\phi_m$, say.[8] If ϕ_m is appreciably greater than ϕ_2 the inelastic scattering increases the single-scattering tail up to ϕ_m by a factor $(1+Z^{-1})$, while the mean deflection, $\alpha_m\delta$, is increased by a factor $(1+Z^{-1})^{\frac{1}{2}}$. If ϕ_m is less than ϕ_2 the inelastic scattering contributes nothing to the single-scattering tail, while the mean deflection is increased by a factor $\{1+\log(M\phi_m^2/\phi_1^2)\div Z\log(M\phi_2^2/\phi_1^2)\}^{\frac{1}{2}}$. In all actual cases the latter factor is practically equal to $(1+Z^{-1})^{\frac{1}{2}}$. Accordingly, apart from the single-scattering tail, the effect of electronic scattering may be allowed for, simply by replacing Z by $(Z^2+Z)^{\frac{1}{2}}$ in the various formulae.

(4)

In the cloud-chamber experiments on the scattering of fast electrons by thin foils the angle of scattering observed is that between the initial direction and the projection of the emergent direction on a plane containing the initial direction and perpendicular to the direction of photography. In Fig. 1 this projected angle of scattering α (assumed small) is represented along the axis OA, the other component, α', of the true scattering angle being represented along OB. The shaded area represents the region in which the scattering is multiple,—with an approximately Gaussian distribution. This area has a radius of nearly $1.5\phi_2$. The question we have to consider here is the effect of a limitation of the component α' to values less than a certain limit ψ, as such limitation operates in the experiments. We shall first consider the case of $\psi \gtrsim 1.5\phi_2$ (ψ_2 in Fig. 1). The scattering in the excluded region is under these conditions practically single scattering, and, accordingly, the probability of a deflection with α between α and $\alpha+d\alpha$, and with $\alpha'>\psi$, is

[8] ϕ_m may be appreciably less than this if a limit is set to the energy lost by the observed electrons.

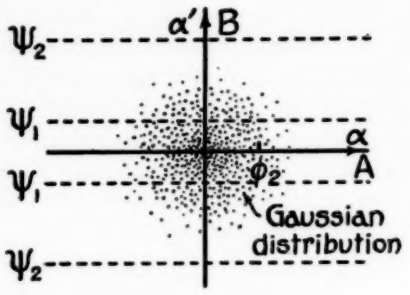


FIG. 1.

(Eq. (ii), p. 563 of earlier paper)

$$P_{ex}(\alpha)d\alpha=(\pi/\alpha^3)\{1-2\beta/\pi-(1/\pi)\sin 2\beta\}d\alpha,$$

where $\tan\beta=\psi/\alpha$. The modified distribution is accordingly

$$P_{in}(\alpha)=P(\alpha)-P_{ex}(\alpha), \qquad (9)$$

where $P(\alpha)$ is the distribution calculated for all α' and given by Eq. (5). Under the conditions concerned, *viz.* $\psi>1.5\phi_2$, the relative value of $P_{ex}(\alpha)$ is very small in the region of multiple scattering, and in this region the distribution is still practically Gaussian. The effect is appreciable only on the single scattering beyond $\alpha\sim\psi$. Integrating the modified distribution it is found that the mean deflection is reduced from the value, $\bar{\alpha}$, given by Eq. (3) to

$$\bar{\alpha}_{in}=(\bar{\alpha}-2\cdot 1/\psi)/(1-2\cdot 2/\psi^2), \qquad (10)$$

all the angles being expressed in terms of δ as unit (e.g., ~ 1).

Next let us consider the case of $\psi<\phi_2$ (ψ_1 in Fig. 1). Under this condition the scattering in the included region up to $\alpha\sim\phi_2$ is nearly Gaussian, so that the distribution of the projected deflection remains Gaussian in shape (as pointed out in the earlier paper). In this region $P(\alpha)$ is everywhere reduced by a constant factor, approximately equal to

$$f_1=(2/\sqrt{\pi})\int_0^{\psi/\sqrt{\pi}\alpha_m} e^{-x^2}dx, \qquad (11)$$

where α_m is given by (6a). For α appreciably greater than ϕ_2 the scattering is single and the distribution of α for the included tracks is

$$(\pi/\alpha^3)-P_{ex}(\alpha)=(\pi/\alpha^3)\{2\beta/\pi+(1/\pi)\sin 2\beta\}=(\pi/\alpha^3)\times f_2. \qquad (12)$$

The change in the *shape* of the distribution is thus equivalent to a reduction of the single-scattering tail (column 3 of Table I) by a factor (f_2/f_1), the mean deflection becoming

$$\bar{\alpha}_{\text{in}}=\left\{\bar{\alpha}-\int_{\phi_2}(\pi/\alpha^3)(1-f_2/f_1)\alpha d\alpha\right\}\Bigg/\left\{1-\int_{\phi_2}(\pi/\alpha^3)(1-f_2/f_1)d\alpha\right\}. \quad (13)$$

The differences between $\bar{\alpha}_{\text{in}}$ and $\bar{\alpha}$, according to both (10) and (13), are small in the actual cases concerned.

For comparison of theory and experiment in the region of multiple scattering a convenient and definite quantity to consider is the mean value of α leaving out all deflections with $\alpha>\phi_2$ (Eq. (5)). We shall denote this mean by $\bar{\alpha}_2$. As we have seen in §(2) the distribution of α is practically Gaussian up to ϕ_2, with a breadth corresponding to a mean deflection, α_m, given by (6a). It follows that

$$\bar{\alpha}_2=\alpha_m(1-e^{-y^2})/\left\{1-(2/\sqrt{\pi})\int_y e^{-z^2}dz\right\},$$
$$y=\phi_2/\sqrt{\pi}\alpha_m. \quad (14)$$

If α' is restricted to $\alpha'<\psi$, and if $\psi<\phi_2$, both α_m and $\bar{\alpha}_2$ remain the same within the accuracy of the results given above. If $\psi\gtrsim 1.5\phi_2$ then for the included tracks

$$\bar{\alpha}_{2,\text{in}}=(\bar{\alpha}_2-J_1/\psi)/(1-J_2/\psi^2), \quad (15)$$

where

$$J_1=\int_{\tan^{-1}(\psi/\phi_2)}^{\pi/2}(\pi-2\beta-\sin 2\beta)\sec^2\beta d\beta,$$

$$J_2=\int_{\tan^{-1}(\psi/\phi_2)}^{\pi/2}(\pi-2\beta-\sin 2\beta)\sec^2\beta\tan\beta d\beta.$$

In actual cases the difference between $\bar{\alpha}_{2,\text{in}}$ and $\bar{\alpha}_2$ is only a few percent, and the corrections are given to show that they are small rather than as quantities to be evaluated accurately (in the present state of the relation of experiment and theory).

In the above discussion it is assumed that the restriction on α' is independent of α, i.e., that ψ is independent of α. An appreciable variation of ψ with α, especially if ψ is small ($<\phi_2$), would lead to serious errors if allowance were not made for it. It is therefore best to have ψ as large as possible, greater than $1.5\phi_2$, and furthermore to have ϕ_2 not greater than about 15° in order to preserve the condition of small angles.

(5) Scattering of Electron Tracks in Gas of Cloud Chamber

In cloud-chamber experiments on the scattering of electrons by foils the incident and emergent directions are obtained from measurements on a finite length, t, of the track on each side of the foil. A certain error is therefore introduced by the scattering which the electrons suffer in the gas. This scattering increases the measured deflections, and its effect may be regarded as an increase in the thickness of the scattering foil. If we denote by α_f the mean deflection produced by the foil, and by α_g the mean deflection produced by a thickness t of the gas, the average fractional increase in the scattering due to the gas is of the order $(\alpha_g^2/\alpha_f^2)\sim(Z_g^2N_gt_g)/(Z_f^2N_ft_f)$, where the suffixes g and f refer to the gas and to the foil, respectively. For lead foils of thickness 0.01 cm, such as have been used, and taking $t=10$ cm of air, $\alpha_g^2/\alpha_f^2\sim 1/400$, so that the scattering in the gas has a negligible effect. However, for an aluminum foil 0.0025 cm thick, as used in one of the experiments of Crane and Slawsky,[9] α_g is comparable with α_f, so that the scattering by the foil is appreciably less than the directly measured scattering. The agreement found between the latter and theory[1] is therefore no longer valid.

It is of interest, particularly in connection with the determination of the energy of a particle from the curvature of its path in a magnetic field, to consider the natural "curvature" of its path due to scattering by the gas. Let $\bar{\theta}$ be the average projected deflection in radians suffered by the particle in a distance t of gas, so that $\bar{\theta}=\bar{\alpha}\delta$ (Eqs. (1) and (3)). Then the uniform radius of curvature, ρ_s, which would give this deflection, is $t/\bar{\theta}$.[10] If the gas is air or a heavier element, and if

[9] H. R. Crane and M. J. Slawsky, Phys. Rev. **56**, 1203 (1939).

[10] The average lateral displacement of the track at the end of the distance t, with respect to the tangent to its initial direction at the beginning of it, may be shown to be $t/\bar{\theta}\sqrt{6}$. The uniform radius of curvature which would give the same displacement is $(3/2)^{\frac{1}{2}}t/\bar{\theta}$.

(t/β^2) is of the order of 10 cm or more, the conditions of multiple scattering are sufficiently well satisfied that $\bar{\alpha}$ is given approximately by Eq. (3). If the gas contains $P\times A$ nuclei per cc where A is Avogadro's number ($\sim 2.7\times 10^{19}$), the average value of ρ_s is, accordingly,

$$\bar{\rho}_s = 0.20\left(\frac{\bar{\beta}t^{\frac{1}{2}}}{\bar{\alpha}ZP^{\frac{1}{2}}}\right)H\rho. \qquad (16)^{11}$$

For fast electrons ($\beta\sim 1$), in air at NTP ($P=2$), and taking $t=10$ cm this gives $\rho/\rho_s = 40/H$. Thus the curvature due to multiple scattering introduces, under these conditions, an error in the determination of the momentum of individual tracks (from their curvature in a magnetic field H) of about 10 percent for $H\sim 400$, and 1 percent for $H\sim 4000$. It is of interest in this connection to consider the curvature due to scattering of the slow mesotron reproduced in a publication by E. Pickup and the writer[12] some time ago. For this mesotron $t\sim 20$ cm, $H=2300$, $\beta=0.4$, effective $P\sim 1$, giving $\rho/\rho_s\sim 1/40$. For a mesotron traversing the last 5 cm or so of its range in air $\beta\sim 0.1$ so that $\rho/\rho_s\sim 500/H$, i.e., a field very large compared with 500 would be necessary to determine its momentum. Actually for such a mesotron $H\sim 3\cdot 10^4$, so that $\rho_s\sim 50$ cm.

It is instructive to compare the radius of curvature, ρ_s, due to scattering with the range of the particle. Let v_0 be the initial velocity of the particle and R its range. Then the average projected deflection of the particle after it has traveled a distance t is

$$\bar{\theta} = \bar{\alpha}\delta = \bar{\alpha}(2Zze^2N^{\frac{1}{2}}t^{\frac{1}{2}}/M\bar{v}^2\xi), \qquad (17)$$

where $\bar{\alpha}$ is given by (3), and $1/\bar{v}^4\xi^2$ is the effective average value of $1/v^4\xi^2$ over the distance t. The nonrelativistic case is the one of greatest interest and for this $\xi=1$, and we may also assume over a limited region of velocity, a power law for the range, i.e., $R\propto v_0^n$. Then it follows that

$$\bar{v}^4/v_0^4 = \frac{4-n}{n}\{f^{-(4-n)/n}-1\}^{-1}\cdot(1-f), \qquad (18)$$

where $f=(R-t)/R$. From the theory of energy loss, the rate of loss of energy is

$$dT/dx = (8\pi NZz^2e^4/mv^2)L, \qquad (19)$$

where $L=\log_e(\sim 137\beta_0 Z^{-\frac{1}{2}})$. This expression for dT/dx corresponds approximately to $n=4-L^{-1}$, and gives

$$R = M_p m v_0^4/8\pi NZz^2e^4(4L-1). \qquad (20)$$

With this value of R we have

$$\rho_s/R = t/\bar{\theta}R = (M/m)^{\frac{1}{2}}Z^{-\frac{1}{2}}F^{-1}, \qquad (21)$$

where $F=\bar{\alpha}\{f^{-(4L-1)^{-1}}-1\}^{\frac{1}{2}}(1-f)^{-1}(2\pi)^{-\frac{1}{2}}$. The quantity F is independent of the mass M_p of the scattered particle, and varies only very slowly with the other quantities involved, *viz.* Z, z and v_0. Taking it as constant we have $\rho_s/R\propto(M_p/Z)^{\frac{1}{2}}$ Thus ρ_s/R is nearly independent of the initial velocity v_0, which means that, for a given Z and M_p, tracks of different initial velocities have the same shape when reduced to the same size. This theorem is true except when the velocity of the particle is comparable with the orbital velocity of the outer atomic electrons, which for protons and α-particles corresponds to a range of half a centimeter or so in air at NTP, and to much less for electrons and mesotrons. For such velocities F ceases to be even approximately constant and the expressions used for R and $\bar{\alpha}$ also fail.

The approximate absolute value of F for velocities, v_0, ranging from about 0.05 c to 0.5 c, and for $f=\frac{1}{2}$, is 0.8, so that

$$\rho_s/R\sim 1.3(M_p/m)^{\frac{1}{2}}Z^{-\frac{1}{2}}. \qquad (21a)$$

This result represents the ratio to the range of the curvature of a track over the first half of its range. For electrons in air it gives $\rho_s/R\sim 0.5$, and this is consistent with the general appearance of electron tracks in air. $\rho_s/R\sim 7$ for mesotrons, ~ 20 for protons, ~ 40 for α-particles. These numbers are, of course, average values, the individual values of (R/ρ_s) having a distribution similar to that of the angle of deflection, that is a Gaussian distribution with a tail due to single scattering.

(6) Multiple Scattering of α-Particles

Owing to the comparative low velocity of α-particles ($\beta=v/c\sim 0.06$), and also their double

[11] Since Eq. (3) refers to the "projected" deflections, Eq. (16) refers to the projected curvature.
[12] E. J. Williams and E. Pickup, Nature **141**, 684 (1938).

electronic charge the quantity

$$\gamma = Zze^2/\hbar v = Zz/137\beta \qquad (22)$$

exceeds unity in most experiments on the scattering of these particles. Under such conditions, as shown in the earlier paper (S, footnote page 537), the use of Born's approximation in calculating the basic single-scattering formula is not justified, and classical mechanics is the valid form of quantum mechanics. This breakdown of Born's approximation, while leaving unchanged the intensity of single scattering by a bare nucleus, increases the angle, ϕ_m, at which the shielding effect of the atomic electrons sets in by a factor approximately equal to γ. This may readily be seen as follows. As long as the distance, p, of closest approach to the nucleus is appreciably less than the radius, a, of the electron cloud surrounding the nucleus the scattering is little affected by the electrons. Their shielding effect sets in when p is of the order of a, and in such collisions the classical momentum transfer is approximately $(Zze^2/a^2)\times(a/v)=Zze^2/av$, giving a deflection $(Zze^2/av)\div M_pv=Zze^2/M_pv^2a$, which therefore represents $\phi_m{}^{cl}$. In Born's first approximation the deflection, $\phi_m{}^B$, produced by the atomic field at distances of the order of a from the nucleus is of the order of $\lambda\!\!\!^{-}/a$ where $\lambda\!\!\!^{-}=\hbar/M_pv$. Thus

$$\phi_m{}^{cl}/\phi_m{}^B \sim (Zze^2/M_pv^2a)/(\hbar/M_pva)=Zze^2/\hbar v=\gamma.$$

ϕ_m enters the formulae for multiple scattering through the quantity M in (3a). M actually represents $\phi_1{}^2/\phi_m{}^2$, where ϕ_1 is such that the particle, in traveling the distance t, suffers on the average one collision giving a projected deflection greater than ϕ_1 (see Eq. (8)). Under the conditions of multiple scattering, i.e., $M\gg 1$, we have $\phi_1\gg\phi_m$, so that the intensity of scattering through angles greater than ϕ_1 is negligibly affected by shielding, and the value of ϕ_1 is the same for $\gamma\gg 1$ as for $\gamma\ll 1$, being determined by the single scattering formula for a bare nucleus (Eq. (8)). Thus under the condition $\gamma\gg 1$ the quantity M is approximately γ^2 times less than the value given in (3a) which is based on the use of Born's approximation. Denoting the former by M_{cl} and the latter by M_B we have

$$M_{cl}=g\gamma^2 M_B=g(2\pi Nth^4/3.1Z^{\frac{1}{3}}m^2e^4)$$
$$=g(2\pi Nta_0{}^2/3.1Z^{\frac{1}{3}}) \qquad (3b)$$

g is a numerical factor of the order of unity, and a_0 is the radius of the hydrogen atom (5.35×10^{-9} cm). In the experiments of Geiger[13] on the multiple scattering of α-particles by gold, for instance, γ^2 is about 500, and the theoretical deflection given by Born's approximation is about 50 percent greater than the true theoretical value given by a classical calculation.

To obtain an accurate estimate of the classical scattering the numerical factor g must be calculated, and this requires a more accurate calculation of the effective minimum angle of scattering $\phi_m{}^{cl}$, in other words a more accurate classical treatment of shielding. This has been done by calculating the classical deflection for different values of the impact parameter, p, assuming a Thomas-Fermi atomic field. In the collisions concerned the absolute deflections are very small and the particle may be assumed to travel uniformly in a straight line when calculating the momentum transfer, $\mu(p)$. The deflection, θ, is then μ/M_pv. Denoting the force acting on the particle normal to its path by $F(x)$, and the distance of the particle from the nucleus in a direction parallel to its path by y we have

$$\mu=\int_{y=-\infty}^{y=+\infty} F(x,p)dy/v=(2Zze^2/pv)f(p/a'),$$

$$f(p/a')=\int_0^1 \{\phi(p/a'(1-z^2)^{\frac{1}{2}})$$
$$-(p/a'(1-z^2)^{\frac{1}{2}})\phi'(p/a'(1-z^2)^{\frac{1}{2}})\}dz.$$

$\phi(u)$ is the ratio of the potential at a point distant ua' from the nucleus to the potential, Zze^2/ua', at the same point if there were no shielding. $a'=0.885a=0.885a_0Z^{-\frac{1}{3}}$. In the absence of shielding $\phi=1$, $\phi'=d\phi/du=0$, and hence $f=1$, so that f represents the ratio of the deflection produced by the atom to that produced by a bare nucleus. $\phi(u)$ is tabulated in Fermi's paper, and f has been calculated from this table for various values of the impact parameter p. The results are given in Table II, the second row giving the values of

TABLE II.

p/a'	128^{-1}	32^{-1}	8^{-1}	4^{-1}	2^{-1}	1	2	4	8	16
	1.00	0.98$_5$	0.95	0.89	0.76	0.60	0.39	0.19	0.08	0.04

[13] Proc. Roy. Soc. **83**, 492 (1910).

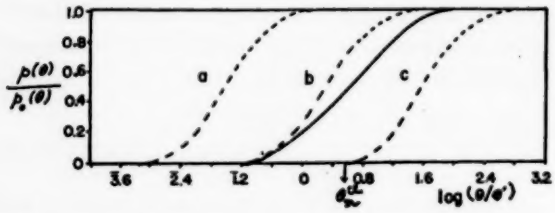


FIG. 2. Effect of shielding on atomic scattering, assuming Thomas-Fermi atomic field. Full curve—classical. Broken curves—Born approximation, (a) for $\gamma=20$, (b) for $\gamma=1$, (c) for $\gamma=0.05$. $\theta'=(Zze^2/M_pv^2a_0Z^{-\frac{1}{3}})$. If $\gamma\gg1$, i.e., if Born-approximation curve is to left of classical curve, the latter is correct representation of quantum mechanics. If $\gamma\ll1$, i.e., if Born-approximation curve is to right of classical curve, the Born approximation is the correct representation.

f corresponding to the values of p/a' in the first row. The effect of shielding on the intensity of scattering, corresponding to the values of f in this table, is shown in Fig. 2 by the full curve. The ordinates represent the ratio of the intensity of scattering, $p(\theta)$, to that due to an unshielded nucleus, $P_0(\theta)$. The abscissae represent log (θ/θ') where $\theta'=Zze^2/M_pv^2a$. The broken curve (a) represents the requirements of Born's approximation for $\gamma=20$ (which is typical of the conditions in experiments on the scattering of α-particles). It will be seen how seriously wrong the approximation is under these conditions. Curves (b) and (c) represent Born's approximation for $\gamma=1$, and $\gamma=0.05$, respectively. For the latter value of γ Born's approximation is of course a valid representation of quantum mechanics. For $\gamma\sim1$ classical theory and Born's approximation give roughly the same results, and neither is as accurate as one or the other is for $\gamma\gg1$ or $\gamma\ll1$. In optical terms the angles of refraction and of "diffraction" of the de Broglie waves by the atomic field are comparable for $\gamma\sim1$, and an exact analysis is difficult. For $\gamma\gg1$ the refraction predominates and diffraction is a small correction. For $\gamma\ll1$ the "diffraction" masks the refraction.

In the problem of multiple scattering shielding enters through the integral

$$\int_0^{\theta_1} P(\theta)\theta^2 d\theta$$

where θ_1 is sufficiently large that $P(\theta_1)$ is unaffected by shielding $(=P_0(\theta))$. This integral has been evaluated numerically* and the effect of shielding corresponds to a sudden cut-off of the scattering at an effective minimum angle $\theta=\theta_m=3.8\theta'$. Actually we require the effective minimum *projected* angle of scattering, ϕ_m. From the earlier paper (S, p 564) $\phi_m=0.83\theta_m$. Thus

$$\phi_m{}^{cl}=3.2(Zze^2/M_pv^2a).$$

The value of ϕ_m using Born's approximation and a Thomas-Fermi atomic field (S, p. 564, Eq. iii) is

$$\phi_m{}^{B}=1.75(\hbar/M_pva).$$

Thus

$$\phi_m{}^{cl}/\phi_m{}^{B}=(3.2/1.75)(Zze^2/\hbar v)=1.8\gamma$$

$$g=1.8^{-2}=0.31$$

$$M_{cl}=0.31(2\pi Nta_0{}^2/3.1Z^{\frac{1}{3}})$$
$$=0.20\pi NtZ^{-\frac{1}{3}}a_0{}^2 \quad (3b)$$

$$M_B=0.64\pi NtZ^{4/3}z^2\beta^{-2}(\hbar/mc)^2. \quad (3a)$$

The multiple scattering is given in terms of M by Eq. (3) irrespective of whether M has its classical value (3b) or the Born-approximation value (3a).

In his experiments on the multiple scattering of α-particles Geiger[13] observed the scattering in space, as distinct from the projected scattering, and he recorded the most probable direction of scattering, Θ_p. The latter falls in the region of Gaussian scattering and in this region the intensity of scattering $P(\Theta)$ is proportional to $\Theta e^{-\Theta^2/\pi\alpha_m{}^2}$, where α_m is the average projected angle of scattering for the Gaussian part of the scattering, and is given in terms of $\bar{\alpha}$ by (6a).

TABLE III.

Scattering Element	Atomic Number	γ	$Nt\times10^{-19}$	M_{cl}	Θ_p in degrees: Observed	Classical	Born Approx.
Gold	79	20	1.08	10	2.1	1.74	3.07
Tin	50	13	1.33	17	1.5	1.43	2.20
Silver	47	12	1.48	20	1.5	1.37	2.11
Copper	29	7	1.91	35	1.1	1.04	1.46
Aluminum	13	3	3.37	105	0.6	0.69	0.85
$\Sigma\Theta_p$					6.8	6.3	9.7

* This was actually done directly from the values of $f(p/a')$ rather than the values of $P(\theta)$ in Fig. 2. In terms of p the integral concerned is $\int_{p_1}^{\infty}\theta^2 p\,dp$, which is proportional to $\int_{p_1}^{\infty}f^2(p/a')d\log p$. p_1 is the impact parameter which gives a deflection θ_1 by an unshielded nucleus.

$P(\Theta)$ is a maximum at

$$\Theta_p = (\pi/2)^{\frac{1}{2}}\alpha_m. \qquad (23)\dagger$$

Geiger gives the values of Θ_p for the scattering of α-particles by foils of different elements, of thicknesses equivalent in stopping power to 1 cm of standard air. The incident α-particles had a range of 5.6 cm corresponding to an incident velocity of 1.80×10^9 cm/sec., an emergent velocity of 1.68×10^9 cm/sec., and an effective average velocity of 1.74×10^9 cm/sec. The number, Nt, of atoms per cm^2 of the foils may be estimated from data on the relative stopping power of different elements (e.g., Rutherford, Chadwick, and Ellis, pp. 100 and 109). Geiger's results are given in the following table together with the theoretical values. The multiple scattering of α-particles has also been investigated by Mayer[14] who observed the projected mean angle of scattering, $\bar{\alpha}$. Following Geiger, Mayer records the scattering by a thickness of foil equivalent in stopping power to 1 cm of air. The α-particles used had a somewhat smaller velocity than in Geiger's experiments, the average velocity being 1.53×10^9 cm/sec. as compared with 1.74×10^9 in Geiger's experiments. The results are given in Table IV. It will be seen that though in certain individual cases in the above tables there are appreciable differences between the observed values and the requirements of theory (represented by the classical values) there is no indication of a systematic departure from theory. Thus while the scattering observed by Mayer for the heavy elements, gold and platinum, is about 25 percent less than the theoretical values, the scattering in gold observed by Geiger is about 20 percent greater than that observed.** The average algebraic difference between the observed and theoretical values is +4 percent in the first table, and −5 percent in the second, differences which are probably well within the errors involved in these early experiments.*

TABLE IV.

Scattering Element	Average projected deflection (degrees)		
	Observed	Classical	Born Approx.
Gold	1.72	2.31	3.36
Platinum	1.85	2.28	3.33
Silver	2.00	1.72	2.33
Copper	1.36	1.26	1.60
Aluminum	0.93	0.81	0.94
$\Sigma\bar{\alpha}$	7.9	8.4	11.6

The theoretical values given by the use of Born's approximation rather than classical theory are also given in the above tables, and are seen to be well removed from the experimental values, being on the average about 40 percent greater. This, to the writer's knowledge, is the first example of atomic collisions which demonstrates the validity of the classical solution of the Coulombian problem under the condition $\gamma\gg1$. In the more familiar problem of stopping power the collisions concerned are between the moving particle (ze) and the atomic electrons (e) and the quantity which determines whether it is classical theory or the Born-approximation that is valid is $\gamma' = ze^2/hv$. Even for α-particles the latter is appreciably less than unity, and it is Bethe's formula for stopping power (Born's approximation) that represents the requirements of quantum mechanics, and which is in agreement with experiment, and not Bohr's classical formula. The difference between Bethe's stopping-power formula ($\gamma'\ll1$) and Bohr's formula ($\gamma'\gg1$) actually corresponds very closely to the difference between the formulae for multiple scattering under the respective conditions $\gamma\ll1$ and $\gamma\gg1$. We shall digress somewhat to show this connection between the two phenomena, and incidentally give a simple derivation of Bohr's and of Bethe's formulae for stopping-power.†

† The scattering in the region of Θ_p is not exactly Gaussian, and the above expression for Θ_p may be in error by 5 percent or so. Actually the more accurate distribution given in the last column of Table I gives a value for Θ_p, for $M=330$, which is about 5 percent less than the value given by (23). Since this section was written a paper by Goudsmit and Saunderson has appeared, in which Θ_p is evaluated for certain cases of fast-electron scattering. Their values confirm those given by (23) within 5 percent, as is shown in Section (8).

[14] Ann. d. Physik **41**, 931 (1913).

** For these heavy elements M is only about 10 and the condition of multiple scattering assumed in the writer's theory, *viz.* $M\gg1$ is not ideally satisfied.

* Geiger and Mayer also observed the multiple scattering by foils of different thicknesses. The writer has not had the opportunity to make a detailed comparison of theory with all the data, but there is no apparent serious discrepancy.

† I am greatly indebted to Professor N. Bohr for the opportunity, during a stay at his institute (1934), of discussing with him the general basis of various collision phenomena, including that of stopping power as outlined here.

The deflection of a particle, mass M_p, through an angle θ by a free electron means the transfer of momentum approximately equal to $M_p v\theta$ to the electron, and hence an energy transfer $Q=(M_p v\theta)^2/2m$. Denoting by $P_e(\theta)d\theta$ the number of collisions per cm in which the particle is deflected through angles between θ and $\theta+d\theta$ by the atomic electrons, assumed to be free, the rate of loss of energy is

$$dT/dx=\int Q(\theta)P_e(\theta)d\theta=(M_p{}^2v^2/2m)\int P_e(\theta)\theta^2 d\theta.$$

The second integral is exactly analogous to that which determines the intensity of multiple scattering. For N atoms per unit volume $P_e(\theta)=8\pi NZz^2e^4/M_p{}^2v^4\theta^3$, giving

$$dT/dx=(4\pi NZz^2e^4/mv^2)\int_{\theta_m'}^{\theta_{\max}} d\theta/\theta$$

$$=(4\pi NZz^2e^4/mv^2)\log(\theta_{\max}/\theta_m').$$

$\theta_{\max}$ is of the order of m/M_p, while θ_m' is an effective lower limit determined by the binding forces. Actually the above expression for $P_e(\theta)$, which applies to free electrons, is in this application valid statistically provided the distance of approach is such that the time of collision is less than the natural period, τ, of the atomic electron, i.e., provided the impact parameter is less than $\rho\sim v\tau/2\pi=v/2\pi\nu$, where ν is the average characteristic frequency of the atomic electron. ρ plays exactly the same role as the shielding radius, a, in multiple scattering, and θ_m' is the deflection of the particle by a free electron when the impact parameter is of the order of ρ. As in multiple scattering the relation between θ_m' and ρ depends on whether $\gamma'\ll 1$ or $\gamma'\gg 1$. If $\gamma'\ll 1$, θ_m' has its value given by Born's approximation, *viz.*, $\hbar/M_p v\rho=h\nu/M_p v^2$ giving

$$(dT/dx)_{\text{Born}}=(4\pi NZz^2e^4/mv^2)$$

$$\times\log\left(\sim\frac{m}{M_p}\frac{M_p v^2}{h\nu}=\frac{mv^2}{h\nu}\right). \quad (24a)$$

If $\gamma'\gg 1$, θ_m' has its classical value, *viz.*, $ze^2/M_p v^2\rho=2\pi ze^2\nu/M_p v^3$ giving

$$(dT/dx)_{cl}=(4\pi NZz^2e^4/mv^2)$$

$$\times\log\left(\sim\frac{m}{M_p}\frac{M_p v^3}{2\pi ze^2\nu}=\frac{mv^3}{2\pi ze^2\nu}=\frac{mv^2}{h\nu}\times\frac{1}{\gamma'}\right). \quad (24b)$$

(24a) is Bethe's formula and (24b) Bohr's formula, and they bear an obvious resemblance to the expressions for $\alpha_1{}^2$ in multiple scattering. The latter under the conditions $\gamma\ll 1$, and $\gamma\gg 1$ are (in radians) (Eqs. (1) and (3a, b))

$$\alpha_1{}^2{}_{\text{Born}}=(4NtZ^2z^2e^4/M_p{}^2v^4)\log M_B,$$

$$\alpha_1{}^2{}_{cl}=(4NtZ^2z^2e^4/M_p{}^2v^4)\log(M_B\times 1/\gamma).$$

Returning to the experiments on α-particles it might be mentioned that observations have been made, particularly by Rose,[15] with foils so thin that the quantity M is less than unity. Under such conditions the scattering remains nearly single for all angles however small. Under these conditions, as the angle of scattering is reduced, the intensity of scattering begins to fall at $\theta\sim\theta_m$ below the intensity of single scattering by unshielded nuclei, *viz.*, $P_0(\theta)$, and at no angle does it materially exceed $P_0(\theta)$. When $M\gg 1$ the scattering, as θ is reduced, first exceeds $P_0(\theta)$ (at $\theta\sim\phi_2$ (Eq. (5)) due to the onset of multiple scattering, and at very small angles again falls below it. Rose's results are typical of $M<1$. The observations are not sufficiently accurate and comprehensive, and the condition $M\ll 1$ not sufficiently satisfied, to make possible a direct quantitative comparison with the classical intensity of single scattering by a shielded nucleus (Fig. 2).

(7) Summary of Results

As the formulae representing the results of the calculations described in this paper and in the writer's earlier paper (S) are somewhat scattered we summarize here the main results.

A natural unit of angle δ is defined by

$$\delta=2Zze^2(Nt)^{\frac{1}{2}}/M_p\beta^2c^2\xi \quad (1)$$

where Z = atomic number, ze, M_p, and βc = charge, mass, and velocity of the scattered particle, respectively, $\xi=(1-\beta^2)^{-\frac{1}{2}}$, N = number of atoms per cc, and t = thickness of scattering material. The small effect of inelastic scattering by the atomic electrons may be allowed for by replacing Z by $(Z^2+Z)^{\frac{1}{2}}$. In terms of δ as unit the arithmetic mean projected deflection is

$$\bar{\alpha}=0.80(\log_e M)^{\frac{1}{2}}+1.45,\ =0.80\bar{\alpha}_1+1.45 \quad (3)$$

[15] Proc. Roy. Soc. **111**, 677 (1926).

where

$$M=M_B=0.64\pi NtZ^{4/3}z^2\beta^{-2}(\hbar/mc)^2, \quad \text{if} \quad \gamma\ll 1, \quad (3a)$$

$$M=M_{cl}=0.20\pi NtZ^{-1}(\hbar^2/mc^2), \quad \text{if} \quad \gamma\gg 1, \quad (3b)$$

$$\gamma=(Zz/\beta)(e^2/\hbar c)=(Zz/137\beta). \quad (22)$$

The probability that the particle has a projected deflection between α and $\alpha+d\alpha$ is

$$P(\alpha)d\alpha=G(\alpha)d\alpha+S(\alpha)d\alpha \quad (4)$$

where

$$G(\alpha)=(1-\pi/2\phi_2^2)(2/\pi\alpha_m)\times\exp[-\alpha^2/\pi\alpha_m^2], \quad (6b)$$

$$\alpha_m=(\bar{\alpha}-\pi/\phi_2)\div(1-\pi/2\phi_2^2), \quad (6a)$$

$$\phi_2=5.1\bar{\alpha}_1-4.0 \quad (5)$$

and

$$S(\alpha)=0 \quad \text{for} \quad \alpha<\phi_2, \quad =\pi/\alpha^3 \quad \text{for} \quad \alpha>\phi_2.$$

The transition from multiple to approximately single scattering takes place in the region of

$$\alpha, \ =\alpha_{ms}, \ =5.1\alpha_m-4.0. \quad (25)$$

For $\alpha>\alpha_{ms}$ the departure from the single-scattering formula, π/α^3, is represented approximately by the factor $(1+3\pi\alpha_m^2/\alpha^2)$. This gives a more definite indication of the extent of multiple scattering than the Wentzel criterion generally used in this connection. The mean projected value of α, excluding α greater than ϕ_2, is

$$\bar{\alpha}_2=\alpha_m(1-e^{-y^2})\div\left\{1-(2/\sqrt{\pi})\int_y^\infty e^{-z^2}dz\right\}$$

where

$$y=\phi_2/\alpha_m\sqrt{\pi}. \quad (14)$$

The most probable deflection in space (as against projected deflection) is

$$\Theta_p=(\pi/2)^{\frac{1}{2}}\alpha_m. \quad (23)$$

For a particle with a range R the average natural curvature, $\overline{\rho_s^{-1}}$, over the first half of its range due to scattering, is given by

$$\overline{R/\rho_s}=F^{-1}Z^{\frac{1}{2}}(M_p/m)^{-\frac{1}{2}} \quad (21)$$

where

$$F=\bar{\alpha}\{2^{(4L-1)^{-1}}-1\}(8\pi)^{-\frac{1}{2}}$$
$$L=\log_e 137\beta_0 Z^{-\frac{1}{2}}$$

and $\bar{\alpha}$ is given by (3) where t, in the expression for M, is equal to $\frac{1}{2}R$. F is nearly a constant and in most actual cases is about 0.8 giving

$$\overline{R/\rho_s}=0.8Z^{\frac{1}{2}}(M_p/m)^{-\frac{1}{2}}. \quad (21a)$$

The essential conditions underlying the applicability of the results is that the average deflection, $\bar{\alpha}\delta$, be small. The finiteness of the deflections gives rise to a fractional increase in $\bar{\alpha}\delta$ approximately equal to $(3\pi/8)(\bar{\alpha}\delta)^2$, and in $\alpha_m\delta$ equal to $(3\pi/8)(\alpha_m\delta)^2$ (appendix C in S, p. 568). For $\bar{\alpha}\delta=10°$ this is about 4 percent. While the estimate of this increase also assumes small angles it is sufficiently accurate to give an indication of the effect of the finiteness of the deflections.

(8) Discussion of Recent Papers*

Two or three papers on the multiple scattering of fast electrons have recently appeared. Sheppard and Fowler[16] have made a further experimental investigation of the scattering of electrons, with energy around 10 Mev, by foils of carbon, aluminum, and lead. They find that the scattering by carbon and aluminum is in fair agreement with the theory given by the writer. However, the scattering by lead is about 40 percent less than the theoretical value. This discrepancy is much greater than the estimated error of 1 or 2 percent in the statistical calculations, and it must be attributed to experimental error or to a failure of the basic single-scattering theory. Actually the statistics (characterized by the value of the quantity M, Eq. (3a)) are practically the same for the carbon and aluminum scatterers as they are for the lead scatterer, so that on this account alone the discrepancy for lead can hardly be attributed to an uncertainty in this part of the theory. The values of M concerned ($\sim$1000) are also intermediate between its values in experiments on the scattering of α-particles ($M\sim 50$) and its values in the experiments on the scattering of cosmic-rays ($M\sim 10^5$), and in both these cases the theory satisfactorily accounts for the observations.

According to Wheeler[17] second-order interference effects may appreciably reduce the scattering, and he concludes that this effect is

* Added in proof (April 12, 1940).
[16] Phys. Rev. **57**, 273 (1940).
[17] Phys. Rev. **57**, 352(A) (1940).

sufficient to account for such a discrepancy as the above for lead. I have not seen details of Wheeler's calculations, but according to the abstract of this paper the effect he considers operates only if the single crystals in the scattering foil exceed a certain critical size, t_c, $\sim(N\lambda\sigma^{\frac{1}{2}})^{-1}$, where N=number of atoms per cc, λ=de Broglie wave-length, σ=scattering cross section. However, for a crystal of this size there cannot be coherence between the secondary waves from all the atoms because of the disturbance produced by inelastic scattering. If an electron ionizes (or excites) an atom situated at P its position with respect to P is thereby determined with an uncertainty not exceeding the "radius of action," ρ, of the electron. Thus the maximum volume of the crystal around P which scatters coherently is defined by the radius ρ.** ρ is the maximum distance from its path at which an electron can ionize. It is determined by the onset of adiabatic conditions, and is equal in order of magnitude to $v(1-v^2/c^2)^{-\frac{1}{2}}/2\pi\nu=v\xi/2\pi\nu$, where $h\nu$ represents the ionization energy. We must now consider the relation of ρ, and the mean free path l for ionization (or excitation) to t_c; for however big the crystal ρ or l, whichever is the greater, limits the effective size of the regions that scatter coherently. For solid material and fast electrons $N\sim(3\times10^{-8})^{-3}\sim(5a_0)^{-3}$, $\lambda=h/mc\xi$, $\sigma^{\frac{1}{2}}\sim a_0Z^{-\frac{1}{2}}(Z/137)$, and $l\sim 137^2\times 8\times 4\pi Na_0^2\times\log(6\times137\xi)$. a_0 is the hydrogen radius. These values give

$$\rho/t_c\sim0.0004Z^{\frac{1}{2}},$$
$$l/t_c\sim0.01Z^{\frac{1}{2}}/\xi(1+0.15\log_{10}\xi).$$

Thus even for the heaviest elements ρ and l are much less than the minimum crystal dimensions assumed by Wheeler, so that the effect considered by him does not take place with electrons.† Apart from these considerations the agreement between the theory and the observed scattering of cosmic-ray particles by lead plates shows that the effect does not exist under such conditions, and this would lead us not to expect it in the case of the scattering by lead foils unless the latter have a very different microcrystalline structure from lead plates.

In view of the discrepancies between the experimental and the theoretical values Goudsmit and Saunderson[18] have considered further the statistics of multiple scattering and have given an alternative treatment of it. A section which they devote to a comparison of their calculations with mine is not very helpful, as it only dwells on details of theoretical procedure without giving any idea of the quantitative relation of the final results, which is the essential consideration, especially to the experimenter. However, a table which they give of the most probable angle of scattering for certain cases investigated experimentally by Crane and Slawsky permits a direct comparison to be made. The most probable angle of scattering, Θ_p, according to my calculations is given by Eq. (23). I have evaluated this for the above cases using the same screening factor as Goudsmit and Saunderson assume, in order that the statistical calculations only be compared. The quantity M (Eq. (3a)) is then equal to 0.27 times the quantity ν tabulated by Goudsmit and Saunderson. Taking the first case in their table, *viz.*, a 0.0025-cm aluminum scatterer, we have $M=0.27\times\nu=26$. $\bar{\alpha}_1=(\log_e M)^{\frac{1}{2}}=1.80$. $\bar{\alpha}=0.80\bar{\alpha}_1+1.45=2.91$. $\phi_2=5.1\bar{\alpha}_1-4.0=5.2$. $\alpha_m=(\bar{\alpha}-\pi/\phi_2)/(1-\pi/2\phi_2^2)=2.43$. $\Theta_p=(\pi/2)^{\frac{1}{2}}\alpha_m=3.04$ (in units δ). $\delta=(2/\sqrt{\pi})\times(\pi K^2Nt)^{\frac{1}{2}}$. G-S also tabulate $(\pi K^2Nt)^{\frac{1}{2}}$, giving $\delta=2.05°$ for the case under consideration. Thus $\Theta_p=3.04\times2.05°=6.2°$. The value obtained by G-S is 6°. The results for all the cases are given in Table V.

It will be seen that the theoretical values calcu-

** The argument could be expressed in an alternative manner in terms of the change in momentum of the electron when it ionizes an atom. The minimum value of the latter is $(h\nu/v\xi)\sim(h/\rho)$, and this leads to a phase-change of the order of $\pi/2$ in a distance ρ.

† Wheeler's effect does not, of course, start suddenly when the crystal size t_c is reached, but for $t\ll t_c$ its quantitative effect on the scattering is negligible. A treatment of the effect of interference between the scattered waves from different atoms could proceed by considering a crystal of dimensions of the order of ρ or l (whichever is greater) as a scattering unit. We would then find that the angular separation of any maxima and minima arising from interference between the secondary waves from different atoms in this unit would be much less than the screening angle ϕ_m, because the distance between the atoms is much greater than the size of the atoms. The intensity averaged over a small solid angle ($\ll\phi_m^2$) would be practically the same as the sum of the intensities in that direction from the different atoms with neglect of interference, thus giving practically the same effective ϕ_m, λ and therefore having a negligible effect on the resultant multiple scattering as the latter is only changed by about 2 percent for a 50 percent change in ϕ_m. This was pointed out in the earlier paper (footnote page 542).

[18] Phys. Rev. **57**, 24 (1940).

TABLE V.

Scatterer	Electron Energy (Mev)	M	δ (degrees)	Θ_p (degrees)	
				Present	Goudsmit and Saunderson
0.0025 Al	1.15	26	2.05	6.2	6
0.010 Al	1.15	104	4.10	14.3	14
0.025 Al	1.15	260	6.50	24.4	23
0.0038 Pb	1.15	250	11.9	44.3	43
0.013 Pb	9.5	760	2.72	11.0	11
0.0063 Pb	4.0	370	4.28	16.4	16
0.3 C	4.0	2,100	4.28	18.1	15

lated by Goudsmit and Saunderson confirm those given by the writer's theory.* The table does not bear out Sheppard and Fowler's statement that G-S obtain smaller values for lead than are given by the writer's theory. It is to be noted that Θ_p refers to the scattering in space and not the projected scattering, the average value of which is dealt with much more accurately in my calculations than the value of Θ_p, the probable error in which is estimated at about 5 percent.

In their reference to the writer's calculations G-S give a somewhat wrong impression of the approximations made, the nature of which we therefore feel necessary to repeat. In the first place, the multiple scattering due to collisions giving deflections less than ϕ_1 is assumed to be Gaussian. The quantitative departure from a Gaussian form was, however, considered in the earlier paper (S) and I have estimated it more carefully since. Provided $M \gg 1$ it is, even for small M, only a few parts in a 1000, so that the assumption is completely justified for all practical purposes. The effect of $\phi > \phi_1$ was not "added as a small correction." The effect increases the average deflection by about 30 percent, and this increase is evaluated numerically using an exact formula (Eq. (V), appendix B in S). The numerical calculations have been carried out for various values of M with an accuracy of 1 percent, and the general formula (3) represents the results of these calculations within about 1 percent (cf. Section (1) of this paper). The effect of $\phi > \phi_1$ on the distribution is to broaden the multiple-scattering part and add a single-scattering tail. The broadened distribution is only approximately Gaussian, and its representation as a Gaussian with an average deflection α_m (Eq. (6)) has not the same degree of accuracy as the representation of the resultant average deflection by $\bar{\alpha}$ (Eq. (3)). The results of the numerical calculation of the distribution in Section (2), however, show that the deviation of the resultant broadened distribution from the Gaussian curve given by (6) is actually very small. This is corroborated by the agreement (Table V) between the values of Θ_p to which Eq. (6) leads and the values given by G-S. The errors of a few percent in the distribution given by (6) are of course of an order of magnitude much smaller than the differences between theory and experiment for lead scatterers.

Besides their reference to the use of Gaussian distributions G-S list three other assumptions or approximations which are usually made. The errors arising from these approximations can, however, be evaluated in order of magnitude and they are very small. One is that the distance traveled by the electron is equal to the thickness of the scatterer. An approximate correction for this assumption was given in the writer's first paper (S, Appendix C, p. 568), and in most of the experiments concerned the whole correction is very small (<3 percent). Disregard of back scattering is another approximation mentioned. However, this is fully justified because in the experiments concerned the thickness of scattering material is chosen to give a small average deflection, and under such conditions both the theoretical and observed probability of back scattering are negligible. Actually the theoretical probability of back-scattering is approximately δ^2 where δ is the unit angle used in the calculations (Eq. (1)). In the experiments of Sheppard and Fowler on fast electrons, for instance, $\delta^2 \sim 0.05^2 \sim 3 \times 10^{-3}$.

Finally G-S list the disregard of inelastic scattering (arising from the emission of radiation) as an approximation. While an exact treatment of the influence of radiative forces on scattering has not yet been given, its approximate absolute magnitude is certainly small. This may be shown as follows. The probability that a nucleus scatters an electron through an angle greater than θ, neglecting radiative reaction, is represented by the cross section $\sigma_s(\theta) = 4\pi Z^2 e^4 / m^2 c^4 \xi^2 \theta^2$. To be

* G. & S. have probably made an arithmetical error in the last case, since Θ_p must be greater than the value of 16° for the preceding case, which has the same δ but a *smaller* M.

deflected through an angle θ through the emission of radiation the energy of the latter must in general be at least of the order of $mc^2\xi^2\theta$. The cross section, σ_r, for the emission of this amount of radiation in a nuclear collision bears to σ_s a ratio of the order of $(4/3\pi)137^{-1}\xi\theta^2 \log (\xi\theta)^{-1} \log \theta^{-1}$, $\sim 3\times 10^{-3}\xi\theta^2 \log (\xi\theta)^{-1} \log \theta^{-1}$. This increases with θ and for the maximum deflection that can in general be produced by radiation, *viz.*, $\theta \sim \xi^{-1}$ (corresponding to emitted energy of the order of the total energy of the electron), we have $\sigma_r/\sigma_s \sim 3.10^{-3}\xi^{-1} \log \xi$. In the experiments on the scattering of fast electrons ξ is of the order of 10, so that $\sigma_r/\sigma_s < 10^{-3}$. We may therefore conclude that the direct effect of the emission of radiation on the scattering is negligible. The indirect effect of inelastic collisions *viz.*, the reduction of the energy of the electrons in the foil is adequately allowed for in the experiments by taking the energy of the electrons as the mean of the incident and emergent energies.

In their treatment of screening Goudsmit and Saunderson use a number of different approximations to the atomic field and state that the results are quite sensitive to the form of field used. If, however, the dimensions of the electron cloud round the nucleus were changed by as much as 50 percent, thus changing ϕ_m by the same amount, the corresponding change in the average deflection (or the width of the multiple-scattering distribution) would, as pointed out in S, be only about 2 or 3 percent. The Thomas-Fermi atomic field, assumed in S, can hardly be in error by more than 10 percent or so, especially for heavy elements, so that from this source we should not expect an error of more than about 0.5 percent in the resultant scattering. A more important source of error in the treatment of screening for heavy elements is the use of the Born approximation, which is valid for fast electrons only if $Z/137$ is appreciably less than unity. This is hardly so for lead. By an interpolation from the classical values of ϕ_m which are valid if $Z/137 \gg 1$ (Section (6), Fig. 2), and the Born-approximation values which are valid for $Z/137 \ll 1$, we infer that the Born-approximation value of ϕ_m is correct for lead within about 20 percent, corresponding to about 1 percent in the resultant scattering.

I wish to take this opportunity of expressing my appreciation of discussions on the problem of scattering with Professor Crane, Professor Goudsmit, and Professor Wheeler during the 1939 Physics Symposium at the University of Michigan.

The Diffraction of X-Rays by Liquid Argon

A. EISENSTEIN AND N. S. GINGRICH
University of Missouri, Columbia, Missouri

(Received May 25, 1940)

The x-ray diffraction pattern of liquid argon has been obtained with crystal-reflected Mo$K\alpha$ radiation. Argon gas under a gauge pressure of 50 lb. per sq. in. was cooled to 90°K in a thin-walled Pyrex glass capillary of 1.7 mm diameter and under these conditions, the argon became liquid. An evacuated camera of 9.53 cm radius was so constructed that the film could be placed outside the vacuum, with entrance and exit windows of aluminum. Exposures of 140 hours were required to give patterns of satisfactory density. Microphotometer records of the films showed three peaks; a very strong peak at $\sin\theta/\lambda=0.154$, a medium peak at $\sin\theta/\lambda=0.280$ and a weak peak at $\sin\theta/\lambda=0.415$. No convincing evidence of another peak could be found beyond this and up to $\sin\theta/\lambda=1.20$. A Fourier analysis of our diffraction pattern results in an atomic distribution curve $[4\pi r^2\rho(r)\ vs.\ r]$ showing approximately 7 atoms at 3.90A and 5 atoms at about 5.05A. A third broad peak occurs at over 7A. This indicates that the distribution of atoms in liquid argon does not follow closely that in crystalline argon, since in the crystal there are 12 nearest neighbors.

INTRODUCTION

THE diffraction of x-rays by liquid argon has been studied by Keesom and de Smedt[1] and their work is of considerable historical interest since it was one of the first investigations of the diffraction of x-rays by liquids. It was probably the first case in which three maxima of intensity were obtained for a liquid. From the positions of these maxima, certain deductions were suggested regarding the structure of liquid argon, but no attempt was made to obtain a complete diffraction pattern. The need for a complete diffraction pattern became evident when the theories of Zernike and Prins[2] and of Debye and Menke[3] were presented. In the present work an attempt has been made to obtain as complete a diffraction pattern of liquid argon as possible, and, from this pattern, to obtain an atomic distribution curve.

EXPERIMENTAL

Since it is essential to use monochromatic x-rays in this work, Mo$K\alpha$ x-rays reflected from a rocksalt crystal were used. A cylindrical brass camera shown in Fig. 1 was constructed so that the distance from the center of the camera to the film was 9.53 cm. A special liquid-air flask was constructed so that it could be waxed in place as shown in the figure. A thin-walled Pyrex glass capillary 1.7 mm in diameter was sealed to the bottom of the flask and connected to a small glass tube leading out near the top of the flask. Argon gas, 99 percent pure, was passed through concentrated H_2SO_4 and then through P_2O_5 to a line providing for evacuation of the capillary tube and for measuring the argon gas pressure. Argon melts at 83.8°K and boils at

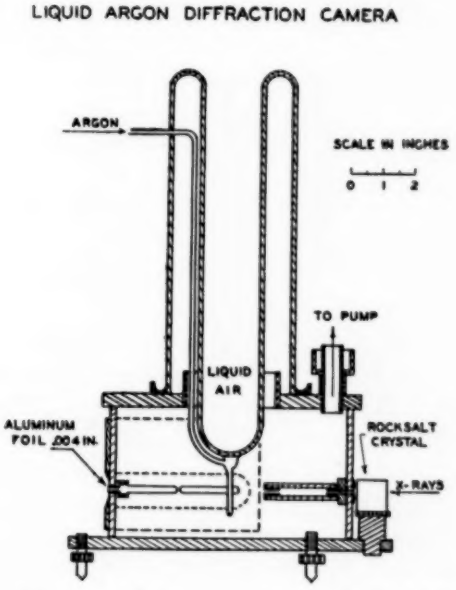


FIG. 1. Diagram of the camera used in obtaining the diffraction pattern of liquid argon.

[1] W. H. Keesom and J. de Smedt, Proc. Amst. Acad. **12**, 1 (1923).
[2] F. Zernike and J. A. Prins, Zeits. f. Physik **41**, 184 (1927).
[3] P. Debye and H. Menke, Erg. d. tech. Röntgenkunde II (1931).

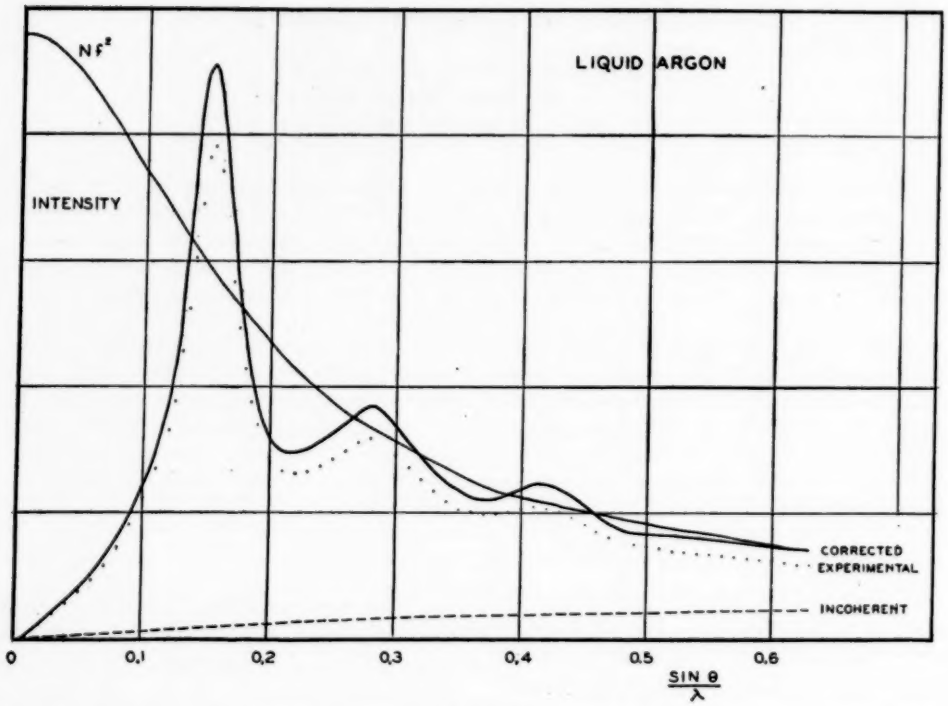


FIG. 2. The experimental and the corrected intensity patterns for liquid argon, taken with monochromatic Mo$K\alpha$ radiation.

87.3°K while oxygen boils at 90°K and nitrogen boils at 77.2°K. Hence liquid nitrogen cannot be used to cool the argon since argon would crystallize at this temperature; nor can liquid oxygen be used since at atmospheric pressure argon would be a gas at this temperature. The procedure adopted consisted of using old liquid air, which was essentially liquid oxygen, and compressing the argon gas to a gauge pressure of 40 to 50 lb./sq. in. The existence of liquid argon in the capillary was verified by measuring the intensity of the x-ray beam leaving the exit window. The inner surface of the liquid-air flask and the outer surface of the capillary were silvered. This thin coating of silver provided a check on the mechanically measured radius since a few weak diffraction lines of silver appeared on the films. The two methods gave values for the radius which checked to within one part in 1000. The camera was evacuated by means of a mercury diffusion pump backed by an oil pump. X-rays were admitted through a small hole covered with 0.004" Al foil and the diffracted x-rays passed through a $\frac{1}{4}$" milled slot covered with 0.004" Al foil. The film, placed outside the vacuum, was held in place by a spring brass door. Several exposures were taken with this camera, though only three were good enough to use for the entire pattern. All three patterns checked well with each other over the range of sin θ/λ reported here. An exposure of 140 hours was required for the best pattern. Because of the construction of this large camera, reliable diffraction patterns could be obtained only as far as about sin $\theta/\lambda=0.7$ or 0.75. In order to investigate the large angle scattering more effectively, an essentially similar camera of 5.08 cm radius was constructed. This made possible the investigation of the diffraction pattern as far as sin $\theta/\lambda=1.20$.

The films so obtained were microphotometered and the intensity pattern is shown in Fig. 2. The experimental curve is shown with dots. There are three distinct peaks located at sin $\theta/\lambda=0.154$, 0.280 and 0.415. No unambiguous and reproducible evidence of a fourth peak could be found. Some indication of a very weak fourth peak appeared on a weak film, but on those films best suited for showing this peak, no definite evidence for its existence could be obtained. Certainly, if

there is one, it must be very much weaker than the third peak. The experimental intensity pattern was corrected and fitted to the f^2 curve in the usual manner,[4] with the resultant curve as shown in Fig. 2. The fully corrected and fitted curve was then analyzed[5] to give the atomic distribution curve, which is shown in Fig. 3.

Discussion of Results

The positions of the intensity peaks obtained in this work can be compared with the early work of Keesom and de Smedt.[1] This comparison can best be made by referring to Table I which lists the peak positions they obtained with Mo filtered and with Cu filtered radiation. Our first peak position happens to fall between their two values, but there is considerable discrepancy in the other two peak positions. Our peak positions were checked on several films and it is believed that of the values listed here, the first peak position is correct to within less than one percent and, because of the uncertainty in locating the highest point on a broad peak, the other two to within possibly two percent.

The atomic distribution curve (Fig. 3) shows a concentration of approximately 7 atoms at 3.90A as the nearest neighbors. The second concentration of about 5 atoms is at 5.05A and the third concentration occurs at somewhat over 7.0A. It is interesting to compare this liquid structure with the crystalline structure of argon. For ready reference, lines whose heights are proportional to the number of neighbors in the crystal are drawn under the atomic distribution curve for the liquid. There are twelve nearest neighbors at 3.84A, six neighbors at 5.43A, twenty-four at 6.64A and so on. The difference in the number of nearest neighbors in the liquid and in the crystal appears to be real. From past experience in this sort of work, it seems that the position of, and the area under, the first peak are relatively unaffected by inaccuracies in the analysis, but that the shape of the first peak is apt to be affected by inaccuracies. Hence we believe this conclusion is based on reasonably good evidence. Similar structural changes have been obtained in the work of Gamertsfelder[6] in the cases of indium and cadmium. Coulson and Rushbrooke[7] have suggested the possibility of structural change on melting in those substances for which there is a large change in density. Argon experiences a decrease of about 11 percent in density in going from the crystal to the liquid phase so that such a change in structure might be expected in this case.

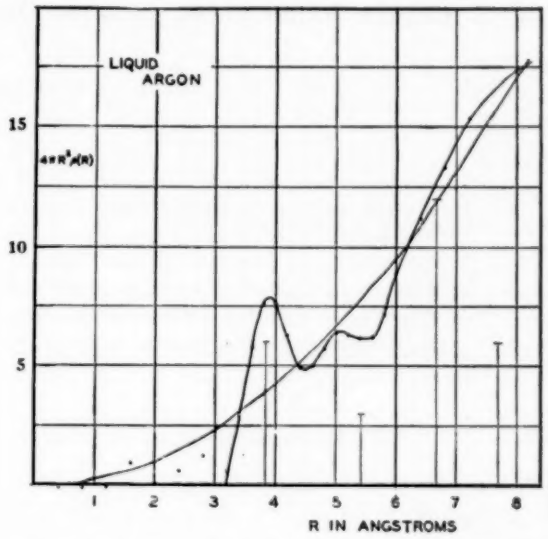


Fig. 3. The atomic distribution in liquid argon. The vertical lines represent the idealized distribution in crystalline argon.

Table I. *Positions of peaks in experimental curves.*

		1	2	3
Keesom and deSmedt				
Mo $K\alpha$	$\sin\theta/\lambda$	0.159	0.235	0.371
Cu $K\alpha$	$\sin\theta/\lambda$	0.151	0.269	—
Present work				
Mo $K\alpha$	$\sin\theta/\lambda$	0.154	0.280	0.415

It is a pleasure to acknowledge grants-in-aid made to one of us (N.S.G.) from the Rumford Fund of the American Academy of Arts and Sciences and from the American Association for the Advancement of Science.

[4] F. H. Trimble and N. S. Gingrich, Phys. Rev. **53**, 278 (1938).
[5] B. E. Warren and N. S. Gingrich, Phys. Rev. **46**, 368 (1934).
[6] C. Gamertsfelder, Phys. Rev. **57**, 1055A (1940).
[7] C. A. Coulson and G. S. Rushbrooke, Phys. Rev. **56**, 1216 (1939).

On the Value of the Electric Quadrupole Moment of the Deuteron*

A. NORDSIECK
Columbia University, New York, New York

(Received May 17, 1940)

The gradient of the electric field at the nucleus in the hydrogen molecule is calculated from a new electronic wave function which gives a simple but accurate formula for the electron density. The wave function is found by a generalization of the procedure first used by S. C. Wang. The calculated field gradient, together with the molecular beam measurements of Kellogg, Rabi, Ramsey and Zacharias, determines the value of the electric quadrupole moment of the deuteron. As stated by the latter authors, the quadrupole moment is 2.73×10^{-27} cm². The writer estimates the limits of error of this value to be ±2 percent.

I. INTRODUCTION

IN a paper on the radiofrequency spectra of the molecules HD and D_2 in a magnetic field, Kellogg, Rabi, Ramsey and Zacharias[1] have pointed out that the evaluation of the electric quadrupole moment of the deuteron requires the knowledge, in addition to the experimentally measured energy differences, of the quantity

$$q=\sum[(3\cos^2\theta-1)/r^3]_{Av}. \qquad (1)$$

The sum is taken over all of the molecular charges except that of the nucleus under discussion, electrons contributing with negative sign. In this expression r is the radius vector from the nucleus under discussion to the individual molecular charge; θ is the angle between r and the direction of the constant applied magnetic field. The average is to be taken over the normal electronic state, the normal vibrational state, and the rotational state J, $M_J=J$, that is, the rotational state for which the projection of the rotational angular momentum on the field direction is a maximum. Casimir[2] has developed the theory of the interaction of a nuclear electric quadrupole moment with extranuclear charges for atoms; except for the actual evaluation of q his theory is immediately applicable to molecules as well.

For an atom the quantity q can be related to another quantity which depends on r in the same way and determines the magnitude of the hyperfine-structure separation; in this way q can be found for the atom from the measured hyperfine structure. In the case of a molecule it is not possible, on account of the lack of spherical symmetry, to make use of such a relation, and we must fall back on a direct calculation of q from the theoretical charge distribution in the molecule. The direct calculation of q for HD and D_2 is the subject of the present paper. A preliminary report of this work was made in February, 1940.[3]

In Eq. (1) q is expressed in terms of coordinates relative to axes with a fixed orientation in space. The theoretical electron distribution, on the other hand, is given relative to axes fixed in and rotating with the molecule. The transformation from the former to the latter axes may be made as follows. Let Θ, Φ be the spherical polar coordinates of the internuclear line relative to the field direction. Let θ' be the angle made by the radius vector r with the internuclear line. Then according to the addition theorem for spherical harmonics

$$3\cos^2\theta-1=(3\cos^2\Theta-1)(\tfrac{3}{2}\cos^2\theta'-\tfrac{1}{2}) \\ +\text{terms involving } e^{i\Phi}. \qquad (2)$$

The average over the rotational state, i.e., over Θ and Φ, may now be found by multiplying by the absolute square of the rotational wave function $|Y_{JJ}(\Theta,\Phi)|^2$ and integrating over Θ and Φ. The terms involving $e^{i\Phi}$ contribute nothing since $|Y_{JJ}|^2$ does not depend upon Φ, and the result is

$$[3\cos^2\theta-1]_{Av}=\frac{-2J}{2J+3}(\tfrac{3}{2}\cos^2\theta'-\tfrac{1}{2}). \qquad (3)$$

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

[1] J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey and J. R. Zacharias, Phys. Rev. **57**, 677 (1940). See pp. 678, 691, 692.

[2] H. B. G. Casimir, "On the interaction between atomic nuclei and electrons," *Teyler's Tweede Genootschap* (1936).

[3] A. Nordsieck, Phys. Rev. **57**, 556A (1940). (The quantity called q in this abstract is called q' in reference 1 and in the present paper.)

Thus if we call the electron density $\rho(r, \theta')$, we have

$$q=\frac{-2J}{2J+3}q', \tag{4}$$

$$q'\equiv\left[\frac{1}{R^3}-\frac{1}{2}\int d\tau\rho(r,\theta')\frac{3\cos^2\theta'-1}{r^3}\right]_{Av}. \tag{5}$$

The relation (4) was stated without derivation in reference 1, page 692. In Eq. (5), R stands for the internuclear distance. The first term is the contribution of the second nucleus; the second comes from the electrons and enters with a negative sign as mentioned above. The average must be taken over the zero-point vibration of the molecule.[4] We shall use the symbol $q'(R)$ for the quantity inside the square bracket.

The integral in (5) is not absolutely convergent, but the theory of a quadrupole moment interaction[2] shows that the integration must be performed in a definite way, which then leads to a unique result in spite of the nonabsolute convergence. One must exclude a small sphere with $r=0$ as center, integrate over the remaining space, and then put the radius of the sphere equal to zero. In practice, therefore, the integration is to be performed in the spherical polar coordinates r and θ', the integration over θ' to be carried out first. This point is mentioned because some proposed electron distributions ρ are not conveniently handled in the coordinates r and θ'.

The physical meaning of q' is the following: $-2eq'$ is the derivative, taken outward along the internuclear line, of the electric field component outward along this line due to all charges outside the small sphere spoken of above.

The chief problem was to find an accurate and convenient approximation for the function ρ. Rather fair accuracy is needed in ρ because of the mutual cancellation of various contributions to q'. The writer has found a simple formula which approximates ρ well for our purposes, by the methods indicated in the following section.

[4] In the definition of q' in reference 1 (end of p. 691 and beginning of p. 692) a factor $\frac{1}{2}$ was accidentally omitted from before the integral sign, and the average over R is not indicated.

II. Calculation of Electron Density

Two approximate electronic wave functions for the normal state were already available for calculating ρ: one found by Wang,[5] another by James and Coolidge.[6] Both were found unsuitable, the first for reasons of accuracy, the second because of its form. The Wang function was found by inserting an effective nuclear charge Z into the first Heitler-London approximation for the wave function and minimizing the electronic energy as a function of Z. The function is very simple:

$$\Psi_{\text{Wang}}=e^{-Z(r_1+r_2')}+e^{-Z(r_1'+r_2)}, \tag{6}$$

r_1 and r_2 being the distances of electrons 1 and 2 from the one nucleus, r_1' and r_2' the distances from the other nucleus, all in atomic units.[7] The electron density is readily found analytically for any value of R. But the density found from this function is in error by as much as 16 percent, as we shall see below. The function of James and Coolidge is in the form of an exponential times a multiple power series with coefficients determined by the Ritz method. It is accurate, but, for our purposes, very unwieldy. In order to calculate q with this function, the density would have to be tabulated numerically for several values of R (a very laborious procedure) and $q'(R)$ found by numerical integration. The fact that the James-Coolidge function is expressed in elliptic coordinates and is not conveniently handled in the coordinates r and θ' contributes to the difficulty. The writer, therefore, decided to use the James-Coolidge function as a standard of accuracy in finding a simple but reliable analytical expression for ρ rather than to calculate with the James-Coolidge function directly.

The errors in the Wang electron density and in the newly found electron density, relative to the James-Coolidge density taken as standard, were estimated as follows. For $R=1.40$ atomic units (the equilibrium distance) the numerical values of the James-Coolidge wave function and of the

[5] S. C. Wang, Phys. Rev. **31**, 579 (1928).

[6] H. M. James and A. S. Coolidge, J. Chem. Phys. **1**, 825 (1933).

[7] Atomic units (D. R. Hartree, Proc. Camb. Phil. Soc. **24**, 89 (1928)), were used throughout in the calculation of q. The value of the atomic unit of length (Bohr radius), upon which the final result for the quadrupole moment depends, was taken to be 0.5292×10^{-8} cm.

wave functions to be tested were computed at 42 chosen points in the configuration space of the two electrons. By squaring these values and averaging, with appropriately chosen weights, over the positions of one of the electrons, comparable values of the electron density were found for the three functions at 9 points in the molecule. These values of the density are called estimated values and are given in the second, third and fifth columns of Table I. In the fourth column are given the errors in the Wang density, in the sixth column those in the new density, both relative to the James-Coolidge density. It should be emphasized that the numbers given are reliable not in an absolute sense but only for purposes of comparison, each wave function having been handled in identically the same way to arrive at the numbers given for it. The values of the density given by the exact formula derived from the new function are also given in the last column, and a comparison of the "estimated" and exact values shows that the points chosen in configuration space and the weights assigned in averaging over the positions of one electron were reasonable.

The table, together with the accompanying scale diagram (Fig. 1) of the positions of the 9 exploratory points in the molecule, shows that the error in the Wang density varies, roughly speaking, in a long range fashion along the length of the molecule, the density being too low in the plane $r=r'$ and too high at the ends of the molecule. If we introduce the elliptic coordinates $\xi=(r+r')/R$ and $\eta=(r-r')/R$ and write the Wang function in terms of these:

$$\Psi_{\text{Wang}}=e^{-\frac{1}{2}A(\xi_1+\xi_2)}\left[e^{\frac{1}{2}A(\eta_1-\eta_2)}+e^{\frac{1}{2}A(\eta_2-\eta_1)}\right], \quad (6')$$

where $A=ZR$, it becomes clear how such a long

TABLE I. *Values of electron density in atomic units (electrons/(Bohr radius)³) at R=1.40 atomic units.*

Point in Molecule (See Fig. 1)	James-Coolidge, Estimated	Wang, Estimated	Error (%)	New Function, Estimated	Error (%)	New Function, Exact Values
a	0.2668	0.2248	−16	0.2694	+1.0	0.2699
b, b'	0.4477	0.4495	+0.4	0.4320	−3.5	0.4304
c, c'	0.06756	0.07734	+14	0.06849	+1.4	0.06825
d	0.1067	0.0942	−12	0.1085	+1.7	0.1087
e, e'	0.04747	0.04686	−1	0.04848	+2.1	0.04849
f	0.00596	0.00606	+2	0.00615	+3.2	0.00616

FIG. 1. The circles $a, b, b' \cdots f$ indicate to scale the points at which the density was estimated. b, b' are the positions of the nuclei. The molecule possesses axial symmetry about the internuclear line. The dotted lines separate the two regions in which the electrons contribute to q' with opposite sign.

range error arises: The trial function has the same coefficient in the η-exponents as in the ξ-exponent, so that the variation of the function in η (variation lengthwise of the molecule) is constrained by being related to the variation in ξ (variation outward from the internuclear line). It is furthermore evident that considerable improvement could be secured while still keeping the very simple analytical form of this trial function, by introducing independent coefficients A and B in the ξ- and η-exponents, respectively:

$$\Psi=e^{-\frac{1}{2}A(\xi_1+\xi_2)}\left[e^{\frac{1}{2}B(\eta_1-\eta_2)}+e^{\frac{1}{2}B(\eta_2-\eta_1)}\right] \quad (7)$$

and minimizing the energy as a function of these two parameters. By carrying out this procedure the new function used in the present calculation was found.[8]

The formula for the energy in terms of A, B and R is given in the appendix, Section 1. The energy can no longer be calculated in completely closed form as was the case with the original Wang function, for the term $1/r_{12}$ representing the mutual potential energy of the two electrons must be expanded by the method of von Neumann[9] and its contribution calculated term by term. Three terms, namely those involving harmonics of the 0th, 1st and 2nd orders, were

[8] Returning to the coordinates r and r', one may interpret this improvement in the wave function from the point of view of the Heitler-London approximation. The present function is formed, not from pure atomic functions, but from atomic functions distorted by the proximity of the second atom. The amount of distortion is measured by $(A-B)$, and is determined by the Ritz variation method.

[9] F. von Neumann, *Theorie des Potentials* (Leipzig, 1887), p. 341.

kept, terms beyond these contributing in order of magnitude 0.0001 atomic unit (0.003 volt) to the energy. Table II gives the calculated minimal values of A and B and the energy for several values of R. For the Wang function at $R=1.400$, $A=B=1.640$ and the electronic energy is -1.851; the James-Coolidge energy at $R=1.400$ is -1.886. The modification of the Wang function has thus reduced the error in the electronic energy from 1.8 percent to 1.2 percent.[10] The formula for the electron density calculated from the new function appears in the appendix, Section 2. It is hoped that this formula together with the estimated errors in it (see Table I) will prove useful in other problems connected with the hydrogen molecule.

III. Calculation of q'

The integrations involved in finding $q'(R)$ are straightforward and lead to the formula given in the appendix, Section 3. To find the value of the part called M_2 it is best to use its expansion in powers of b/a, which is also given. The values of $q'(R)$ are given in Table III.

The average value q' of $q'(R)$ over the zero-point vibration of the molecule, depends somewhat on the amount and character of the anharmonicity in the molecular potential curve. The average was calculated by fitting a formula $a/R+b/R^2+c/R^3$ to the values in Table III, fitting a Morse potential curve to the vibrational data for the molecule,[11] and averaging over the lowest vibrational state in this potential. This approximate procedure is justified because the amplitude of the zero-point vibration is a small fraction of the equilibrium distance between nuclei (0.035 for HD, 0.028 for D_2). The correction due to centrifugal force in the rotational state $J=1$ is negligible (0.2 percent). For the molecule HD the averaged q' was found to be 0.1768 atomic units $=1.193\times10^{24}$ cm^{-3}.[7] For D_2, which has a slightly smaller amplitude of vibration, $q'=0.1763$ atomic units $=1.190\times10^{24}$ cm^{-3}. It is accidental that the average q' is nearly equal to the value of $q'(R)$ at the equilibrium distance, for the form of $q'(R)$ favors smaller values of R, while the anharmonicity favors larger R.

Table II. *Data for new electronic function, all in atomic units. Energy does not include nuclear repulsive term.*

R	A	B	Energy
1.300	1.613	1.265	−1.915
1.400	1.706	1.344	−1.863
1.500	1.793	1.420	−1.815

[10] The root-mean-square error of any function relative to the James-Coolidge function is easily estimated by the exploratory method described above. For the Wang function we find 17 percent; for the new function 12 percent. These are of the order (relative error in energy)$^{\frac{1}{2}}$, as is to be expected on the basis of general theorems. From the present work it appears, however, that provided long range errors in the trial function are sufficiently reduced, the mean relative error in the *density* is of the same order as that in the energy.

[11] C. R. Jeppeson, Phys. Rev. **45**, 480 (1934); **49**, 797 (1936).

IV. Estimate of Accuracy of q'

The accuracy of the values of $q'(R)$ was estimated as follows: For $R=1.400$ the integrand for $q'(R)$ was integrated separately over the two angular ranges separated by the dotted lines in the figure. The two integrands in r (corresponding to the contributions to $q'(R)$ of opposite sign) were then plotted and estimates made of the errors in these integrands at several points, based on the errors in ρ given in Table I. In this way it was found that the errors in ρ at the points a, c and d caused an error of roughly $+1$ percent in $q'(R)$, while the error at the point b' caused an error of roughly -1 percent in $q'(R)$. The other errors had negligible effect because of their being weighted with $1/r^3$. The contribution of the electrons to $q'(R)$ is very nearly $-\frac{1}{2}$ the contribution $1/R^3$ of the second nucleus, so that the fractional error in $q'(R)$ is equal to the fractional error in the electron contribution. Since these estimates were necessarily rough, limits of error of ±1 percent were assigned to $q'(R)$. The process of averaging over the zero-point vibration may lead to an additional error of at most ±1 percent in the average q'. Hence the writer considers 2 percent a safe estimate of the limits of error of the final value of q'. Roughly the same limits of error then hold for the value[1] 2.73×10^{-27} cm^2 of the quadrupole moment, since the experimentally measured term differences are reliable to about $\frac{1}{2}$ percent.

Table III. *Values of $q'(R)$ in atomic units.*

R	$q'(R)$
1.300	0.2462
1.400	0.1755
1.500	0.1257

Appendix

1. The electronic energy in atomic units, not including the energy of mutual repulsion of the nuclei, as calculated from the approximate function (7), is given by

$$E=\left(1-\frac{k}{S^2+T^2}\right)\left(\frac{A}{R}\right)^2-\frac{l+m_0+m_1+m_2+\cdots}{S^2+T^2}\frac{A}{R},$$

where:

$$S_1=\left(\frac{1}{2}A+\frac{1}{2}A^2\right)e^{-A}, \qquad S_1'=\left(\frac{1}{2}A-\frac{1}{2}A^2\right)e^{A},$$

$$S=\left(1+A+\frac{1}{3}A^2\right)e^{-A}, \qquad S'=\left(1-A+\frac{1}{3}A^2\right)e^{A},$$

$$S_3=\left(\frac{3}{A}+3+\frac{3}{2}A+\frac{1}{2}A^2\right)e^{-A}, \qquad S_3'=\left(\frac{3}{A}-3+\frac{3}{2}A-\frac{1}{2}A^2\right)e^{A},$$

$$S_4=\left(\frac{12}{A^2}+\frac{12}{A}+\frac{17}{3}+\frac{5}{3}A+\frac{1}{3}A^2\right)e^{-A}, \qquad S_4'=\left(\frac{12}{A^2}-\frac{12}{A}+\frac{17}{3}-\frac{5}{3}A+\frac{1}{3}A^2\right)e^{A},$$

$$y_0=\sinh B/B; \quad y_n=(\partial/\partial B)^n y_0,$$

$$T=y_0\left(1+A+\frac{1}{2}A^2\right)e^{-A}-y_2\left(\frac{1}{2}A^2\right)e^{-A},$$

$$T'=y_0\left(1-A+\frac{1}{2}A^2\right)e^{A}-y_2\left(\frac{1}{2}A^2\right)e^{A},$$

$$k=\left[\frac{1}{3}S(A^2+B^2)+\frac{1}{2}(y_0-y_2)T(A^2-B^2)\right]e^{-A},$$

$$l=4(S+y_0T)(1+A)e^{-A},$$

$$m_0=-(1/A)\Big\{(S^2+T^2)\log\gamma A+2(SS'+TT')E(2A)-(S'^2+T'^2)E(4A)$$
$$-\left[\left(-\frac{5}{8}A+\frac{3}{4}A^2+\frac{1}{3}A^3\right)+y_0^2\left(-\frac{5}{8}A+\frac{3}{4}A^2+\frac{1}{2}A^3\right)-y_0y_2\left(\frac{1}{2}A^3\right)\right]e^{-2A}\Big\},$$

$$m_1=+\frac{3}{A}\Big\{y_1^2\left[S_3^2\log\gamma A+2S_3S_3'E(2A)-S_3'^2E(4A)-\left(18+\frac{95}{8}A+\frac{35}{4}A^2+\frac{5}{2}A^3+\frac{1}{2}A^4\right)e^{-2A}\right]$$
$$-2y_1y_3\left[S_1S_3\log\gamma A+(S_1'S_3+S_1S_3')E(2A)-S_1'S_3'E(4A)-\left(3A^2+\frac{5}{4}A^3+\frac{1}{2}A^4\right)e^{-2A}\right]$$
$$+y_3^2\left[S_1^2\log\gamma A+2S_1S_1'E(2A)-S_1'^2E(4A)-\frac{1}{2}A^4e^{-2A}\right]\Big\},$$

$$m_2=-\frac{45}{16A}\Big\{(3y_2-y_0)^2\Big[S_4^2\log\gamma A+2S_4S_4'E(2A)-S_4'^2E(4A)$$
$$-\Big(\frac{288}{A^2}+\frac{192}{A}+128+\frac{871}{18}A+\frac{43}{3}A^2+\frac{26}{9}A^3+\frac{1}{3}A^4\Big)e^{-2A}\Big]-2(3y_2-y_0)(3y_4-y_2)$$
$$\times\Big[SS_4\log\gamma A+(SS_4'+S'S_4)E(2A)-S'S_4'E(4A)-\Big(24+16A+\frac{22}{3}A^2+\frac{19}{9}A^3+\frac{1}{3}A^4\Big)e^{-2A}\Big]$$
$$+\Big[\Big(\frac{4}{15}\Big)^2+(3y_4-y_2)^2\Big]\Big[S^2\log\gamma A+2SS'E(2A)-S'^2E(4A)-\Big(2A^2+\frac{4}{3}A^3+\frac{1}{3}A^4\Big)e^{-2A}\Big]\Big\},$$

γ = Euler–Mascheroni constant = 0.5772,

$$E(x)\equiv\int_x^\infty e^{-y}\frac{dy}{y}.$$

2. The electron density in atomic units derived from the approximate function (7) is

$$\rho(\xi,\eta)=\frac{2}{\pi}\Big(\frac{A}{R}\Big)^3\frac{1}{S^2+T^2}e^{-A\xi}[S+T\cosh B\eta]$$

or

$$\rho(r,r')=\frac{2}{\pi}\Big(\frac{A}{R}\Big)^3\frac{1}{S^2+T^2}e^{-A(r+r')/R}\Big[S+T\cosh\frac{B}{R}(r-r')\Big].$$

A and B for various R are given in Table II. S and T are defined in Appendix 1.

3. $q'(R)$, as calculated from the above formula for the electron density, is given by

$$q'(R)=\frac{1}{R^3}\Big[1-\frac{LS+MT}{S^2+T^2}\Big],$$

where

$$L=(24A+12A^2+4A^3/3)e^{-A}-12S\log 2\gamma A-12S'E(2A),$$

$$M=M_1+M_2,$$

$$M_1=2A^3\Big\{\Big[\frac{4}{a^3}-\frac{5b}{a^3}-\frac{15b^2}{a^5}-\frac{b^2}{a^3}\Big]e^{-b}-\Big[\Big(\frac{4}{a^3}+\frac{4}{a^2}+\frac{2}{a}+\frac{2}{3}\Big)-\Big(\frac{15}{a^5}+\frac{15}{a^4}+\frac{6}{a^3}+\frac{1}{a^2}\Big)b^2\Big]e^{-a}$$
$$+\Big[\Big(\frac{9}{2a^4}+\frac{9}{2a^3}+\frac{2}{a^2}+\frac{1}{2a}\Big)b-\Big(\frac{15}{2a^6}+\frac{15}{2a^5}+\frac{3}{a^4}+\frac{1}{2a^3}\Big)b^3\Big][\log(a+b)-\log|a-b|+\overline{\mathrm{Ei}}(a-b)]e^{-a}$$
$$+\Big[\Big(\frac{9}{2a^4}-\frac{9}{2a^3}+\frac{2}{a^2}-\frac{1}{2a}\Big)b-\Big(\frac{15}{2a^6}-\frac{15}{2a^5}+\frac{3}{a^4}-\frac{1}{2a^3}\Big)b^3\Big]E(a+b)e^a\Big\}.$$

M_2 = same as M_1 with a and b interchanged and $\overline{\mathrm{Ei}}(a-b)$ replaced by $-E(a-b)$.

$$M_2=A^3b\Big\{\frac{4}{15}\Big[\frac{1}{a^2}-\Big(\frac{1}{a^2}+\frac{1}{a}\Big)e^{-a}+e^{-a}-aE(a)\Big]$$
$$-\frac{4}{35}\Big[\frac{6}{a^4}-\Big(\frac{6}{a^4}+\frac{6}{a^3}+\frac{3}{a^2}+\frac{1}{a}\Big)e^{-a}+\Big(\frac{1}{3}-\frac{a}{6}+\frac{a^2}{6}\Big)e^{-a}-\frac{a^3}{6}E(a)\Big]\Big\}+0(b^3),$$

$$a\equiv A+B;\quad b\equiv A-B.$$

$\overline{\mathrm{Ei}}(x)$ is defined in Jahnke-Emde, *Funktionentafeln* (Leipzig, 1933), p. 79.

Temperature Effects in Secondary Emission

D. E. WOOLDRIDGE
Bell Telephone Laboratories, New York, New York

(Received June 20, 1940)

Measurements have been made on the effects of temperature changes on the emission of secondary electrons from iron, nickel, cobalt, and molybdenum. Abrupt changes of one or two percent were observed to accompany the $\alpha-\gamma$ transition of iron, while the hexagonal to face-centered cubic transformation of cobalt was accompanied by a change in secondary emission of only about 0.4 percent. The magnetic transformation was found to alter the secondary emission coefficient of nickel by less than 0.3 percent. The temperature coefficient of secondary emission, in the cases of nickel, cobalt, and molybdenum was found to be much less than the volume coefficient of expansion of the metal. The smallness of the temperature coefficient and the effect of the magnetic transformation are shown to lend support to the view that the secondary electrons are scattered or "absorbed" by an excitation process similar to that whereby they are originally produced.

THE nature of the dependence of secondary emission on temperature is of some theoretical importance. The work of Treloar and Landon[1] on iron, nickel and cobalt has shown that temperature plays no very great role, but their methods were incapable of detecting changes of less than about 5 percent in the secondary emission of their targets. This paper deals with more precise measurements of the effects of temperature on the secondary emission of iron, nickel, cobalt and molybdenum.

Apparatus and Technique

There is one secondary emission datum that can be accurately determined more easily than any other. This is the bombarding voltage (V_1) for which the secondary emission coefficient (δ) is exactly unity. For, when $\delta=1$, there is no net flow of current to or from the target; this is true regardless of what bombarding current is used. It is easy to devise a circuit[2] which will enable the observer to set the net flow of current to the target sufficiently near to zero that the accuracy of the determination of V_1 is limited by the precision with which the voltage can be read,—and the measured V_1 does not depend on the primary current.

In all the work to be described V_1 has been measured as a function of temperature. The data can easily be translated to variation in δ at constant V provided that we know the slope of the δ *vs.* V curve near $\delta=1$.

In the work done with iron, the target consisted of a 0.004″ diameter wire, in a tube similar to that used by Treloar[3] in measurements on filaments. The target wire was heated by passing a current through it. The temperature was roughly determined by measurement of the resistance of the wire, supplemented by appropriate calculations of end cooling and radiation effects.

In all the rest of the work, the target consisted of a disk of thickness 0.015″ to 0.020″ and diameter 0.500″. This disk was confronted by the open end of a molybdenum cylinder, which was maintained at a positive potential high enough to insure collection of all secondary electrons. The primary beam came through a mesh-covered hole in the opposite end of the cylinder and covered a spot of about $\frac{3}{16}$″ diam. on the target. The target was heated by thermal radiation from a tungsten filament close behind the disk, and temperatures were measured by means of a thermocouple junction of fine Pt and Pt Rh wires welded to one edge of the target.

In all tubes, of course, pure, pre-glowed tungsten filaments were used as the source of primary current, and were operated at low enough temperatures to make evaporation of tungsten entirely negligible. The amount of metal used in the tubes was kept to a minimum. The parts were heat-treated before assembly,

[1] L. R. G. Treloar and D. H. Landon, Proc. Phys. Soc. **50**, 625 (1938).
[2] In the present work a one-stage d.c. amplifier was used.
[3] L. R. G. Treloar, Proc. Phys. Soc. **48**, 488 (1936).

and the designs were such that all metal parts could be heated in the assembled tubes by induction or bombardment. After the tube was baked and the elements glowed the target was heat-treated. Iron and molybdenum were outgassed by heating for a few minutes at an elevated temperature (1000° for iron, 1170° for molybdenum); nickel and cobalt targets were held at 850° for a few minutes, and then were kept at about 750° for twelve hours or more.

The resulting pressure, as measured with an ionization gauge, was always less than 10^{-7} mm Hg, which was also an upper limit for the pressures reached at any time during the subsequent measurements. The measurements were made on the pumps, but only in the case of molybdenum was there any indication of gaseous adsorption. In this case, measurements had to be made rapidly—an entire heating or cooling cycle lasting only a few minutes—in order to avoid difficulties with the adsorption of gas at the lower temperatures.

Experimental Results

Iron

Figure 1 shows three typical V_1 *vs.* T curves plotted from measurements on a pure iron wire. The temperature scale is far from accurate in this work on filaments. Every plot of V_1 against filament heating current showed a break at a value of current which, from the calculations, had to be near 900°. Hence it was thought safe to assign to the average of these current values the temperature corresponding to the α to γ transition of iron—910°C. The remainder of the temperature scale undoubtedly decreases in accuracy as T varies away from this value.

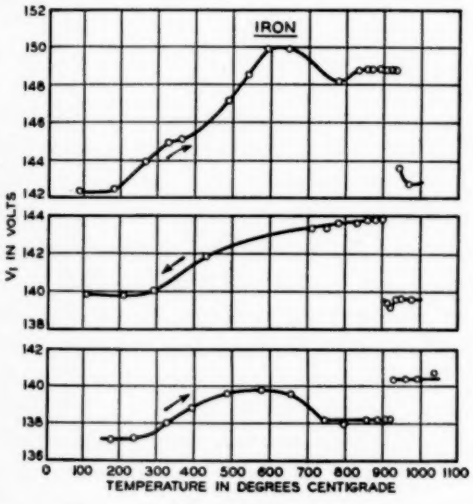


FIG. 1. Effect of crystal structure on the secondary emission from a small iron wire (one percent change in $V_1 = 0.6$ percent change in δ).

In spite of the fact that all three curves of Fig. 1 correspond to measurements made in the same way on the same target, the results are

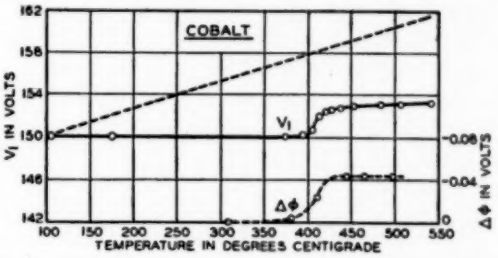


FIG. 2. Effect of crystal structure and temperature on the secondary emission parameter (V_1) and work function (ϕ) of a polycrystalline disk of cobalt (one percent change in $V_1 = 0.2$ percent change in δ), dotted line representing a temperature coefficient of secondary emission equal to the volume coefficient of expansion.

seen to be quite at variance with one another. The curves agree only in that, in every case, an abrupt change in V_1 occurs when the crystal form changes from body- to face-centered, or vice versa. The magnitude and even the direction of this change vary from run to run, as does the behavior of V_1 at lower temperatures. The explanation is probably about as follows. The heat treatment to which the iron wire was subjected resulted in the formation of long crystals, so that the small region in the center of the filament on which measurements were made consisted of only one or two single crystals of iron. (Subsequent photomicrographs of the iron filament have confirmed this.) Consequently each transition through 910° caused a particular crystal surface of α-iron to be replaced by some γ-iron surface configuration. It is not to be expected that exactly the same crystal orientation would be repeated on cooling and reheating. Thus the differences in the results of successive measurements correspond to differences in the crystal surfaces exposed to the primary bombardment. The large, continuous changes in V_1 at temperatures below 910° were probably due partly to continued recrystallization effects and partly

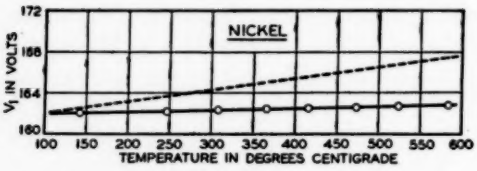


FIG. 3. Effect of magnetic state and temperature on the secondary emission from nickel (one percent change in $V_1 = 0.5$ percent change in δ), dotted line representing a temperature coefficient of secondary emission equal to the volume coefficient of expansion.

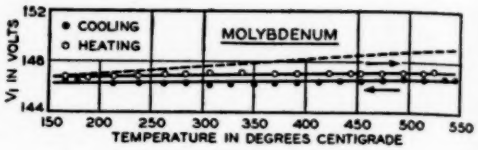


FIG. 4. Effect of temperature on the secondary emission from molybdenum (one percent change in $V_1 = 0.4$ percent change in δ), dotted line representing a temperature coefficient of secondary emission equal to the volume coefficient of expansion.

to shifting of the bombarding spot along the crystals of the target due to thermal expansion and contraction of the filament.

Thus the measurements on iron filaments tell us that the secondary emission coefficient of some surfaces of a body-centered iron crystal differs by at least one or two percent[4] from the secondary emission coefficient of some of the face-centered crystals, but that the variation from surface to surface of a body-centered crystal can be greater than this.

Cobalt

Figure 2 shows the results of measurements on a disk of pure[5] cobalt. In this, and all subsequently described work, the bombarded area was large enough to cover many crystals. There can be little doubt that the abrupt change in V_1 at 410° corresponds to the close-packed hexagonal to face-centered cubic crystal structure change which has been reported[6] to occur at a number of temperatures ranging between 400° and 490°C. The data of Fig. 2 were taken for ascending temperatures, and were thoroughly reproducible, only a few minutes being required to establish equilibrium even for a measurement around 410°. On cooling, the phase change appeared to occur much more slowly, however, and, unless a number of hours were consumed in making the measurements, the decrease in V_1 was distributed over 100° or more.

With this target the contact potential difference between the clean tungsten filament and the cobalt surface was measured simultaneously with the measurements of V_1. The variation in the work function of the cobalt is plotted as the dotted curve of Fig. 2. Besides being too small, the change in work function is in the wrong direction to be directly responsible for the change in secondary emission that accompanies the phase change.

The broken line has the slope that the V_1 *vs.* T curve would have if δ varied with T as rapidly as the volume of the target metal changes with temperature. The slope of this line involves in an obvious way the coefficient of cubical expansion and the slope of the δ *vs.* V curve for cobalt near $\delta = 1$. The latter figure has been taken from measurements made by Treloar and Landon.[1]

Nickel

Figure 3 shows the dependence of the secondary emission of pure[7] nickel on temperature. It is notable that there is evident no influence produced by the magnetic transformation (Curie point = 355°C).

For the broken line showing the comparison with the volume expansion of the metal, the slope of the δ *vs.* V curve is taken from measurements made by the author on solid nickel.

Molybdenum

Figure 4 shows the dependence of the secondary emission of stock molybdenum on temperature. As has been mentioned earlier, the apparent ease with which molybdenum absorbs gas made it necessary for these measurements to be made very rapidly. The two sets of data of Fig. 4 were taken for successive cooling and heating, each set of data requiring about five minutes, with a similar interval elapsing between the cooling and the following heating run.

For the line showing the comparison with the volume expansion of the metal, the slope of the δ *vs.* V curve is taken from Warnecke.[8]

[4] For converting from variations in V_1 to equivalent changes in δ, the slope of the δ *vs.* V curve for iron was taken from R. Kollath, Physik. Zeits. **38**, 202 (1937).

[5] Impurities <0.1 percent.

[6] For a bibliography see L. Marick, Phys. Rev. **49**, 831 (1936).

[7] Impurities <0.2 percent.

[8] R. Warnecke, J. de phys. et rad. **7**, 270 (1936).

Theory and Discussion of Results

The theoretical considerations involved in a prediction of the effects of temperature on secondary emission are similar to those discussed in an earlier article.[9] In that paper particular attention was paid to the secondary electrons originating in the highest occupied energy band of the metal. The nearly quantitative agreement of some of the resulting predictions with experiment prompted the suggestion that excitation of electrons from the lower bands was perhaps not important. However, it can be shown that qualitatively similar results would be predicted by the same type of theory even though some of the lower bands do contribute to the process. Hence it will be of value to place the present considerations, if possible, on somewhat more general principles than some of those employed in the earlier work.

Even though the electrons dealt with are not nearly free, it can be shown (by a slight modification of the earlier theory) that the only probable interactions with primary particles of not too low energy are those which transfer the lattice electrons to levels in higher bands which have the same reduced wave vectors as the initially occupied levels. When the primary energy is not too low, the number of electrons per unit primary path length knocked from occupied levels of reduced wave vector $\sim\mathbf{k}$ in band B into corresponding, initially unoccupied, levels in band B' is

$$(dn/dx)_{\mathbf{k};\,B\to B'} = 4\pi e^4 \rho_B(\mathbf{k}) |b_{\mathbf{k};\,B\to B'}|^2 / |\Delta E_{\mathbf{k};\,B\to B'}|^2. \quad (1)$$

$\rho_B(\mathbf{k})$ is the density of occupied levels of reduced wave vector $\sim\mathbf{k}$ in the lower band, $\Delta E_{\mathbf{k};\,B\to B'}$ is the energy difference between the initial and final states, while

$$|b_{\mathbf{k};\,B\to B'}|^2 = \frac{1}{\Omega^2}\left|\sum_{\mathbf{m}}\left\{\left[\int \psi_{\mathbf{k};B} \exp\left[-i\mathbf{k}\cdot\mathbf{r}-i(2\pi\mathbf{m}/a)\cdot\mathbf{r}\right]d\tau_r\right]\right.\right.$$
$$\left.\left.\times\left[\int \psi^*_{\mathbf{k};\,B'} \exp\left[i\mathbf{k}\cdot\mathbf{r}+i(2\pi\mathbf{m}/a)\cdot\mathbf{r}\right]d\tau_r\right]\right\}\right|^2. \quad (2)$$

Here the two ψ's represent the probability amplitudes for the lattice electron in the initial and final states, normalized so that

$$\int \psi\psi^* d\tau_r = \Omega. \quad (3)$$

Ω, of course, is the volume dealt with. The equation assumes a simple cubic lattice structure, of edge a, but the results are substantially the same for other structures. In terms of a simple cubic structure, $\mathbf{m}$ is to be summed over all vectors having integral components along each of the three cubic axes.

Equation (1) differs only slightly from Eq. (29) of the earlier paper, and is derived in practically the same way.

Temperature coefficient

The influence of temperature on secondary emission is partially determined by the volume dependence of $|b|^2$ and $|\Delta E|^2$. For a material in which the atoms are so far apart that the electrons are tightly bound, neither $|b|^2$ nor $|\Delta E|^2$ depends on a. For a material in which the electrons are nearly free, $|\Delta E|^2 \propto 1/a^4$,[9] whereas the volume dependence of $|b|^2$ varies with $\mathbf{k}$. For metals such as those we are concerned with in this paper, the average $|\Delta E|^2$ will probably depend on a somewhat less rapidly than through the fourth power, while the average $|b|^2$ will decrease with increasing volume at a rate not greater than three or four powers of a. In other words, the quotient $|b|^2/|\Delta E|^2$ will probably increase with a not much more rapidly than a^3, whereas the dependence may be considerably less than this.

The secondary emission coefficient of a target is proportional to the product of the number of secondaries produced per unit primary path, the reciprocal of the absorption or scattering coefficient of the secondaries, and a term which expresses the fraction of the secondaries which are initially directed so as to be able to leave the target surface, if not absorbed. In the nearly free electron case, this last term (within a numerical factor) is

$$[1-\{(W_a)/(E_0+E_F)\}^{\frac{1}{2}}],$$

where W_a is the height of the potential barrier

[9] D. E. Wooldridge, Phys. Rev. **56**, 562 (1939).

at the target surface, and E_0+E_F is the average kinetic energy with which the secondaries are produced. Both W_a and E_0+E_F should decrease with increasing a, at nearly enough the same rates that the temperature dependence of this term, which could not be large in any case, can be ignored. This should be true enough in the more general case also.

The volume dependence of the absorption coefficient of the secondaries, γ_s, depends upon the mechanism by which the secondaries are prevented from escaping. Two types of "absorption" may be important. They may best be discussed in connection with a concrete example. Consider the case of nickel, which possesses overlapping, incompletely filled $3d$ and $4s$ bands, with higher reaching unfilled $4p$ bands. The secondary electrons that escape probably are first excited into $4p$ levels. On their way to the target surface they may excite other electrons from $3d$ to unfilled $4s$ levels, losing large increments of energy in the process, or they may raise some of the electrons at the top of the filled portions of the $3d$ and $4s$ bands to slightly higher levels, losing small increments of energy in this process. The second type of energy loss has been shown[9] to be unimportant as regards the more energetic primary particles. If we assume the first process to be of controlling importance in slowing down the secondaries also, then the expression for γ_s must be similar in form to Eq. (1), involving, probably, nearly the same volume dependence. The net result would be a secondary emission coefficient nearly independent of temperature expansion.

If, on the other hand, the secondaries were scattered by small energy transfers to electrons in the upper levels of the occupied bands, the volume dependence of γ_s would probably not be the same as that of Eq. (1). The net result would be a secondary emission coefficient with a temperature dependence which might be as large as or larger than the cubical coefficient of expansion.

The observed smallness of the temperature coefficient of secondary emission therefore suggests that the mechanism which scatters the secondary electrons is an excitation process similar to that which originally produces them.

Magnetic transformation

The effect of the magnetic transformation should also be small. The influence on the secondary emission should probably be due principally to the change in the fractional occupancy of the $4s$ band and the shift in the energies of the occupied $3d$ and $4s$ levels, where the secondaries originate. Neither of these effects can be large; one can estimate (very roughly) that the resulting influence on the secondary emission would be of the order of several tenths of one percent. The experimental observations on nickel set an upper limit of 0.2 to 0.3 percent upon the effect on δ of the magnetic transformation. As in the case of the temperature coefficient of secondary emission, the experimental result seems lower than one would expect, but not enough lower to be conclusive. Once again, the low value is more easily understood if we assume the scattering of the secondaries to be altered by the magnetic transformation in much the same way that the creation of secondaries by the primaries is affected.

Crystal structure change

A change in the crystal structure of a target entails changes in $|b|^2$, $|\Delta E|^2$ and γ_s. In general, we would expect to observe differences of the same order of magnitude among the various faces of a single crystal. The measurements on iron bear this out. The values of δ measured on what we have inferred to be different crystalline surfaces of the iron wire differ among themselves by only two or three percent. The theory calls for the existence of crystalline directions along which the emission of secondaries is considerably higher than in other directions.[9] Even though γ_s depends on direction in much the same way, it seems unlikely that the two effects can compensate so well that all crystal surfaces have values of δ that agree this closely. It is probable that a complete set of measurements on all the crystal surfaces of iron would reveal values of δ involving differences of at least 10 percent—possibly more.

The effect of the crystal-structure change in cobalt was even less than in iron. In this case, however, microscopic examination of the target (after use, of course) showed that the grain size was so small that several thousand crystallites

had been included in the surface area on which measurements were made. If these crystals were randomly oriented, the result of a structure change would be dependent on changes in the mean values of $|b|^2$, $|\Delta E|^2$, and γ_s, appropriately averaged over all directions. For the cobalt transformation, changes in the physical properties are very slight, and it is likely that the average $b^2/|\Delta E|^2\gamma_s$, which determines δ, would be only negligibly affected. If, however, as is often the case in sheet metal, preferred orientation existed in the target, we must suppose that it was preserved in the higher temperature phase in such a way as to leave nearly unaltered the average orientation of the directions of high emission.

Mr. C. D. Hartman has been of great assistance in making the measurements.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

Scattering of Thermal Neutrons by Crystals

FRANCO RASETTI
Laval University, Quebec, Canada

(Received June 3, 1940)

The effect of crystal interference on the scattering coefficient for thermal neutrons was investigated under different conditions. Calcite shows a large interference effect, the effective molecular cross section for scattering by a single crystal being about three times smaller than for a microcrystalline aggregate. It is also shown that a single crystal becomes more transparent when the velocity of the neutrons is decreased. Scattering has also been measured for bismuth and lead, where the difference between a single crystal and the microcrystalline material is less pronounced. The scattering in the latter case is the same as found for the liquid. Cooling the crystal to liquid nitrogen temperature does not produce any considerable change in scattering. The cross section for 90 percent Pb^{206} was measured and found to be the same as for the ordinary isotope mixture.

INTRODUCTION

RECENT experiments have shown that the scattering coefficient of a substance for thermal neutrons is not generally equal to the product of the scattering cross section per nucleus by the number of nuclei per unit volume. Particularly in the case of crystalline substances,[1] interference between the neutron de Broglie waves scattered by the atoms in the lattice may give considerably large effects, mostly in the sense of making the total "effective" cross section for a molecule of a compound smaller than it would be expected by adding the cross sections of the component atoms.

The aim of the present investigation was to study these phenomena more systematically than it had hitherto been done, in order to establish the relative importance of different factors which determine the interference effects. The main points taken into consideration were: comparison of scattering in the solid and liquid state; dependence of the scattering coefficient on the size and perfection of the crystal, on its temperature, and on the velocity of the neutrons.

EXPERIMENTAL PROCEDURE

The experimental set-up essentially consisted of a slow neutron source of the "howitzer" type, and a BF_3 ionization chamber connected to a linear amplifier.

The neutron source was a stainless steel tube, 14 mm in diameter and 25 mm in length, filled with a mixture of 250 mg Ra (as sulfate) and 1.25 g Be powder.[2]

The well in the paraffin "howitzer" had a

[1] M. D. Whitaker and H. G. Beyer, Phys. Rev. **55** 1101 (1939); **55**, 1124 (1939); M. D. Whitaker, W. C. Bright and E. J. Murphy, Phys. Rev. **57**, 551 (1940).

[2] The author wishes to acknowledge the cooperation of Eldorado Gold Mines, Ltd. in preparing a source which proved satisfactory from the standpoint of neutron yield and compact geometry.

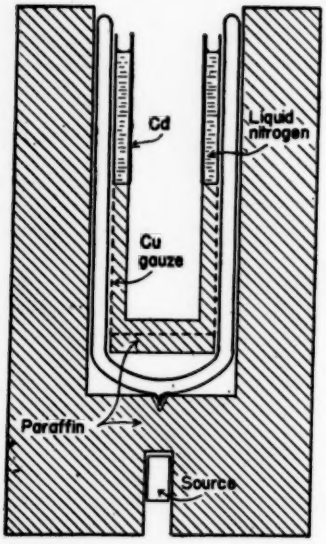


FIG. 1. Arrangement for production of low temperature neutrons.

diameter of 5 cm and a depth of 20 cm, the source being placed 3 cm below the bottom of the well. In most of the experiments, a B_4C diaphragm limited the slow neutron beam used to a diameter of 32 mm. The BF_3 chamber was cylindrical, 2.5 cm in diameter and 14 cm in effective length; the shape was chosen in order to combine fairly high sensitivity with good geometry in the scattering experiments. The chamber was filled at atmospheric pressure, and heavily shielded with Cd and B_4C on all sides, except the front. The distance between the front of the chamber and the opening of the howitzer varied between 5 and 15 cm.

Further diaphragms of Cd and B_4C were used to limit the effective size of the scatterers when necessary.

The ionization kicks were recorded by means of a linear amplifier of conventional design, a thyratron scaling system, and a Cenco impulse meter. The thyratron system was biased so that only the large kicks due to the boron disintegration were recorded. The number of impulses per minute without scatterer was of the order of 250. A Cd sheet of 0.5 g/cm² reduced the number to about 10, and a thick B_4C layer to about one or two per minute. It therefore appears that all but 4 percent of the neutrons recorded have thermal energies. In all data here reported, it must be understood that the background due to neutrons transmitted by Cd has been measured and subtracted.

The source of low temperature neutrons consisted of a cylindrical soft glass Dewar flask placed inside a large paraffin block, the Ra+Be source being located outside the Dewar itself. A paraffin cylinder was placed inside the Dewar, and cooled by liquid nitrogen poured into a ring-shaped metal container. The set-up is illustrated in Fig. 1. A copper gauze extending through the paraffin layer speeded up considerably the cooling process, a thermocouple in the paraffin indicating practically liquid N_2 temperature 30 minutes after the beginning of the operation. The liquid N_2 consumption to maintain the temperature was very low.

As in former experiments[3] it was found that the neutrons had not attained thermal equilibrium with the paraffin, their average velocity (measured from the absorption coefficient in B) being only 1.50 times lower than at room temperature.

Throughout the experiments, the sensitivity of the apparatus remained remarkably constant, because of careful stabilization of all voltage supplies. This fact, coupled with the constancy of the source, enabled one to reproduce a result, at a distance of several weeks, within the accuracy of one or two percent.

RESULTS AND DISCUSSION

Scattering by calcite

A systematic investigation of the scattering of thermal neutrons by calcite was undertaken. This material appeared especially appropriate for studying interference effects, on account of (*a*) the facility of securing large crystal specimens, (*b*) the particularly perfect structure of the lattice, (*c*) the lack of elements possessing any considerable absorption, (*d*) the presence of practically one isotope for each element, (*e*) the practical absence of nuclei of odd mass number, which excludes the complications due to spin effects in the neutron-nucleus interaction.

The cross sections for scattering in carbon[4] and oxygen[5] have been measured fairly accurately.

[3] G. A. Fink, Phys. Rev. **50**, 738 (1936); O. R. Frisch, H. v. Halban, and J. Koch, Proc. Danske Vidensk. Selskab **15**, 10 (1938).

[4] M. Goldhaber and G. H. Briggs, Proc. Roy. Soc. **A162**, 127 (1937).

[5] H. Carroll and J. R. Dunning, Phys. Rev. **54**, 541 (1938).

The cross section of calcium was redetermined with cylindrical scatterers of Ca metal, up to 5.4 g/cm², since the value given by Dunning and co-workers[6] might easily be suspected of being too high, having been determined from the oxide. The value found for the total cross section is actually much lower, i.e., $4.4\pm0.2\times10^{-24}$ cm². The capture cross section is probably only a small fraction of the total.

Assuming then the following values for the cross sections (in 10^{-24} cm²): oxygen 4.1; carbon 4.5; calcium 4.4; one obtains, adding the nuclear cross sections and neglecting interference effects, a total cross section for the $CaCO_3$ molecule

$$\sigma_{\text{add}}=21.2 \text{ (in } 10^{-24} \text{ cm}^2\text{)}. \quad (1)$$

The experimental value for calcite varies quite considerably according to the characteristics of the material employed. The scattering was first measured in a perfect specimen of optical Iceland spar 10.0 g/cm² thick, the neutron beam being normal to the faces of the rhombohedra. This showed a mass scattering coefficient 0.0395 ± 0.0010 cm²/g, corresponding to a molecular cross section

$$\sigma'=6.6. \quad (2)$$

The large difference between values (1) and (2) indicates that interference effects are actually very important in calcite, as found for quartz by Whitaker and Beyer.[1] They also observed that a single crystal of quartz is about twice as transparent as the powdered or amorphous material.

The reason of this fact is easy to understand.[7] Assume that part of the scattered intensity is distributed over a certain number of "Laue spots" produced by Bragg reflection of the neutrons on internal lattice planes of the crystal;[8] whereas another part of the scattering is incoherent, and consequently not selective with respect to direction. A single crystal will then readily eliminate from the neutron beam those neutrons whose directions and velocities are such as to fulfill the Bragg condition for some lattice plane of the crystal. The process will take place in a thin layer of the crystal, and due to the sharp selectivity of the Bragg reflection, will affect only a very small fraction of the incident neutrons. Further layers will contribute to the weakening of the beam only through incoherent scattering. The effective cross section determined for a perfect single crystal will therefore essentially give a measure of the incoherent scattering. In a powder with random orientation, instead, both coherent and incoherent scattering will contribute to the weakening of the beam for any thickness of the scatterer, and consequently the effective cross section will be larger.

The following experiment was performed. Several large cleavages of the same variety of transparent (but not optically perfect) calcite were chosen. Measurements performed on single crystals showed a mass scattering coefficient of 0.047 ± 0.002, the scattering being practically exponential up to thicknesses of 20 g/cm² (this value is higher than the one reported above, probably due to imperfections of the crystals). The material was ground, and samples containing cleavages of fairly uniform size were separated by means of sieves. Absorption cells of aluminum foil were filled with the powders of different sizes, each cell containing 5.0 g/cm² of the material. Table I gives the effective mass scattering coefficient as function of the average number n of crystals traversed by a neutron in the scatterer.

The last value in the table was, however, obtained from precipitated calcium carbonate. The powder was heated for one day at 300° C, then, while still hot, compressed into a metal cylinder and immediately sealed. Notwithstanding these precautions, the possibility that the value of the scattering coefficient might be too high on account of water adsorbed in the powder must not be overlooked. The molecular cross section obtained from this last value is slightly in excess of the additive value (1).

TABLE I. *Scattering in calcite as function of crystal size.*

n	MASS SCATTERING COEFFICIENT μ IN CM²/G
1	0.047 ± 0.002
13	0.070 ± 0.003
39	0.079 ± 0.003
150	0.098 ± 0.003
1.5×10^4	0.139 ± 0.005

[6] J. R. Dunning, G. B. Pegram, G. A. Fink and D. P. Mitchell, Phys. Rev. **48**, 265 (1935).
[7] G. C. Wick, Physik. Zeits. **38**, 403, 689 (1937); O. Halpern, M. Hamermesh and M. H. Johnson, Phys. Rev. **55**, 1125 (1939).
[8] Bragg reflection of neutrons by crystals has been directly observed by D. P. Mitchell and P. N. Powers, Phys. Rev. **50**, 1205 (1936).

Table I supplies a rough evaluation of the percentage of neutrons scattered by a single crystal through Laue reflection. Let us assume that each crystal scatters a certain fraction k of the incident neutrons by Laue reflections (independent of thickness, provided the crystals are not too small), and let the mass scattering coefficient of a thick single crystal be μ. Then, the apparent scattering coefficient μ' of a powder containing an average number n of crystals in a thickness δ will be given by:

$$\mu'\delta=\mu\delta+nk. \qquad (3)$$

From the figures $\mu'=0.070$ cm²/g, $\mu=0.047$ cm²/g, $\delta=5$ g/cm², $n=13$, one obtains that a single crystal scatters about one percent of the incident neutrons through Laue reflections. This figure could hardly be compared with the theoretical width of the reflections, since the observed value is essentially determined by the mosaic structure of the crystal, as for the case of x-rays.

By trying to apply formula (3) to the other data of Table I, one would obtain smaller values of k. The probable explanation is that, with decreasing size of the crystals, before the Laue reflections cover the whole spectrum of the incident neutrons, each crystal does no longer completely scatter even within the region of selective reflection. It appears that the limiting value of the effective cross section for very small crystal size approximately coincides with the additive value.

Dependence of scattering on neutron energy

The relative transparency of single crystals due to interference effects should be more pronounced for neutrons of lower velocity, which have a longer de Broglie wave-length. This was proved by the two following experiments.

The neutron beam emerging from the room temperature howitzer was filtered through a calcite single crystal of 8.9 g/cm² thickness. A Pyrex plate of 0.40 g/cm² was found to transmit 48 percent of the calcite filtered neutrons and 53 percent of the original neutrons, indicating a slightly lower average velocity of the filtered neutrons.

A check was provided by measuring the scattering coefficient of the same calcite single crystal for room temperature neutrons ($\mu=0.047\pm0.002$) and for "cold" neutrons ($\mu=0.038\pm0.002$). The effect is in the same sense, that is, the neutrons of lower average velocity are less scattered. It is to be noticed that any true absorption by nuclei would tend to give an effect of the opposite sign.

Theoretical considerations by Wick[7] indicate that a crystal should be practically transparent to neutrons of a wave-length $\lambda>2a$, where a is the side of the elementary lattice cell, provided that certain conditions be fulfilled (temperature low compared with the characteristic Debye temperature; absence of isotopes and of nuclear spin; absence of true absorption). No crystal seems to satisfy all of these requirements and have a sufficiently small lattice constant (except possibly diamond) to make the effect observable with the neutron source available.

Scattering by lead and bismuth

In most crystalline materials, the effect of interference on the effective scattering cross section does not appear to be as large as in quartz or calcite.

In order to analyze the different factors which may play a part in the phenomenon, other substances were investigated. These were chosen such as to make possible a comparison of the scattering coefficient in the solid and liquid state. Lead and bismuth appeared indicated for the experiment.

Two identical cylindrical cans of thin steel contained each 30.8 g/cm² of lead. In one the lead was in form of a solid cylinder, in the other it was kept liquid by means of an electric heater. The two scatterers could be alternatively placed across the neutron beam. The solid sample had been hammered in order to make the size of the crystals as small as possible. Etching with nitric acid showed that their average linear dimensions were of the order of $\frac{1}{2}$ mm.

The mass scattering coefficient was found to be 0.027 ± 0.002 for the liquid and 0.028 ± 0.002 for the solid. Thus, within the limit of experimental error, no difference appeared between the effective cross sections in the liquid and solid state. Since in the former case interference effects are likely to be negligible, the corresponding value $\sigma=9.5\times10^{-24}$ cm² will represent the true value of the nuclear cross section.

However, a considerable difference is found between a single crystal and the microcrystalline metal. A single crystal of lead, 20.3 g/cm² in thickness, was grown by slow cooling in an electric oven. Etching showed that the whole block actually consisted of one crystal. Its scattering coefficient was carefully compared with that of a microcrystalline (hammered) sample, also 20.3 g/cm² thick.

The result was:

Pb single crystal, $\mu=0.0213\pm0.001$,
Pb microcrystalline, $\mu=0.0275\pm0.001$.

Among the reasons which make the effect so much smaller in the present case than in calcite, the following may be suggested as important: (*a*) The mixture of isotopes in the case of lead; if the different isotopes have cross sections of different magnitude, a diffuse scattering will be superimposed upon the regular Laue reflection. (*b*) If, as probable, the neutron-nucleus interaction depends on spin orientation, the same effect as in case (*a*) will be due to the random orientation of the nuclear spins in the lattice. (*c*) The Debye temperature of lead is far below room temperature; oscillations of the lattice are therefore strongly excited and this fact may contribute some additional diffuse scattering.

Some information on the relative importance of these factors may be gathered from the following results. The same experiments were repeated with bismuth, which is quite similar to lead as far as most of the relevant properties are concerned, except that it consists of one isotope, Bi^{209}. A single crystal, 20.9 g/cm² thick, was compared with the same thickness of the liquid metal, the result being:

Bi single crystal, $\mu=0.019\pm0.001$ cm²/g,
Bi liquid, $\mu=0.024\pm0.002$ cm²/g.

The difference, although significant, is no larger than in lead, and very much smaller than in calcite between the single crystal and the microcrystalline form. In the present case, either the nuclear spin or the oscillations of the lattice would appear mainly responsible for the high percentage of incoherent scattering.

To separate the two last factors, a bismuth single crystal was placed in contact with a steel cylinder which could be filled with liquid nitrogen. The crystal, contained in an evacuated tube with convenient thin glass windows, was placed across the neutron beam, and the transmitted intensity was measured by alternatively keeping the Bi at room temperature and at 80°K. The values found are:

Bi single crystal at 300°K, $\mu=0.0184\pm0.0005$,
Bi single crystal at 80°K, $\mu=0.0171\pm0.0005$.

The difference, if any, is quite small. This fact suggests that the much higher percentage of incoherent scattering observed in Bi than in calcite may be mainly due to the nuclear spin.

A comparison between ordinary lead and radiogenic lead extracted from the Great Bear Lake pitchblende was also attempted. According to the mass-spectrum analysis of Nier,[9] the latter consists of almost 90 percent Pb^{206}. This fact would lead one to expect a lower percentage of diffuse scattering from a crystal of the radiogenic lead than from ordinary lead, due to the much smaller effect of isotopes and of nuclear spin (Pb^{206} having $I=0$).

Unfortunately, repeated attempts to obtain a single crystal of the radiogenic lead completely failed, the same procedure which easily worked with ordinary lead giving a microcrystalline aggregate. The comparison could only be made under the latter form. Within the limits of the experimental error (about 5 percent) no difference was found between the scattering coefficient of ordinary lead and of 90 percent Pb^{206}.

[9] A. O. Nier, Phys. Rev. **55**, 153 (1939).

Spin-Orbit Coupling in He^5

S. M. DANCOFF
Department of Physics, University of California, Berkeley, California

(Received May 27, 1940)

Experimental results on the splitting of the ground state of He^5 are compared with the structures which would be expected on the basis of (*a*) Thomas relativistic spin-orbit coupling, and (*b*) the tensor spin-orbit interaction of mesotron theory. It is found that the Thomas coupling, as customarily formulated, gives an inverted doublet which is several orders of magnitude too small. The tensor force gives a result of the same order of magnitude as the experimental splitting and of the "normal" structure; these latter results can be established, however, only for a nonsingular behavior of the potential function. We discuss a modified treatment of the relativistic spin-orbit coupling, suggested by Breit, which would account for the existence of widely spaced spin-orbit multiplets.

I

RECENT experiments of Staub[1] on the anomalous scattering of neutrons in helium have set the energy of the resonant level at about 1 Mev and have determined the structure to be that of a doublet with a separation of 0.25 ± 0.1 Mev. In a Hartree model, one would suppose the ground state of He^5 to consist of a P neutron outside an alpha-particle core; such a configuration would, if spin-orbit forces were present, give rise to a doublet with J values $\frac{1}{2}$ and $\frac{3}{2}$. The relative intensities of the observed peaks agree fairly well with the ratio to be expected for such a doublet (2.5 to 1 in the forward direction); for this, however, it is necessary to assume that one has the "inverted" order, the $\frac{3}{2}$ level lying lower.

It does not seem possible to fit the observations by other assumptions regarding the low-lying states of He^5. Tyrrell[2] has made calculations which indicate that a state with the extra neutron in a higher S orbit would lie very close to the above P state. If one were to assume that the observations do not indicate a spin-orbit splitting, but rather resonance for both P and S states, the P state lying slightly lower, one would obtain an intensity ratio of 9 to 1 in the forward direction, far outside the range of experimental error.

One is then led to try to explain the magnitude and direction of the splitting on the basis of the effect of proposed nuclear spin-orbit forces on a ^2P state.

II

The Thomas relativistic spin-orbit force yields directly an inverted order of the doublet.[3] If the corresponding perturbing term in the Hamiltonian is taken as $\hbar\boldsymbol{\sigma}\cdot[\dot{\mathbf{v}}\times\mathbf{v}]/2c^2$, the magnitude of the splitting can be estimated very simply. If we adopt the Hartree model the effect is due entirely to the precession of the P neutron. We consider it to be acted upon by a central force represented by a "square well" potential of depth V_0 and radius a. The acceleration is then $-\mathbf{i}(1/M)(\partial V/\partial r)$,[4] and the term becomes

$$-\frac{\hbar^2}{M^2c^2}\left(\frac{1}{r}\frac{\partial V}{\partial r}\right)(\boldsymbol{\sigma}\cdot\mathbf{L}), \tag{1}$$

where $\hbar\mathbf{L}=M[\mathbf{r}\times\mathbf{v}]$. The factor $(\boldsymbol{\sigma}\cdot\mathbf{L})$ introduces a coefficient $+1$ for $J=\frac{3}{2}$ and -2 for $J=\frac{1}{2}$. The splitting is

$$-\frac{3}{4}\frac{\hbar^2}{M^2c^2}\int r^2dr\left\{\frac{1}{r}\frac{\partial V}{\partial r}P^2(r)\right\}. \tag{2}$$

Here $P(r)$ is the radial part of the neutron's

[1] The author is indebted to Dr. Staub for communication of his results before publication. H. Staub and H. Tatel, Phys. Rev. **57**, 936 (1940). *Note added in proof.*—A more complete analysis of Dr. Staub's indicates that by taking into account the interference of s and p scattering, the data can be fitted rather better on the assumption of a normal than an inverted doublet, with the splitting still about 250 kev. This is indirectly confirmed by scattering of 2-Mev neutrons where it is reported that there exists a preponderance of small-angle scattering (low energy recoils). This would favor a normal doublet and, for 180-degree scattering as in Staub's case, destructive interference between s and p scattering. Thus, at present, the evidence, though still inconclusive, seems to indicate a normal doublet such as would be expected, as this paper shows, from known nuclear forces.

[2] W. A. Tyrrell, Jr. Phys. Rev. **56**, 250 (1939).

[3] D. R. Inglis, Phys. Rev. **50**, 183 (1936).

[4] We will use throughout $\mathbf{i}=\mathbf{r}/r$.

wave function; we take $P(r)\sim re^{-r/\lambda}$ and obtain

$$-\frac{\hbar^2}{M^2c^2}\frac{a^3}{\lambda^5}V_0e^{-2a/\lambda}. \quad (3)$$

A relation between V_0 and a is obtained by setting up the condition for a resonant state of the P neutron near ground. This is

$$V_0=(\pi^2/2)\hbar^2/Ma^2. \quad (4)$$

A fairly good estimate of a, the radius of the potential trough, can be obtained from Staub's experimental data. The cross section for the spherically symmetric nonresonance scattering should approximately satisfy the relation $\sigma_S=4\pi a^2$. This gives $a=4.6\times10^{-13}$ cm. This formula for the S wave cross section is derived for a high repulsive barrier such as is customarily used in calculating "potential" scattering. Although there is in this case a Fermi repulsion at short distances between the incident neutron and the neutrons in the α-particle, the model is probably not sufficiently trustworthy to establish accurately the value of a. On the other hand, the value obtained is consistent with the expectation that the radius of the trough in He^5 should be considerably larger than the radius of the α-particle; for if the incident neutron can remain bound for any length of time it must do so by transferring most of its kinetic energy to the other particles, which consequently take up orbits of larger radii.† The quantity λ is more uncertain, but it should not be appreciably smaller than a. With $a=\lambda=4.6\times10^{-13}$ cm, the splitting is

$$-\frac{\pi^2}{2}\frac{\hbar^4}{M^3c^2}\frac{a}{\lambda^5}e^{-2a/\lambda}\sim2.8 \text{ kev}. \quad (5)$$

This result is too small by two orders of magnitude to fit the observed splitting; it seems very improbable that the uncertainties involved in the estimate could be responsible for so large a discrepancy.

Another method of estimating the Thomas splitting consists in using the centripetal acceleration, v^2/r, in place of $(1/M)(\partial V/\partial r)$. This gives, instead of (1):

$$-\frac{\hbar^4}{4M^3c^2}\left(\frac{1}{r^4}\right)(\boldsymbol{\sigma}\cdot\mathbf{L})L^2. \quad (7)$$

The eigenvalue of L^2 is here $L(L+1)=2$. Proceeding as before, the splitting is

$$-\frac{3}{2}\frac{\hbar^4}{M^3c^2}\int r^2dr\left(\frac{1}{r^4}\right)P^2(r)=-\frac{\hbar^4}{M^3c^2\lambda^4}, \quad (8)$$

which, with the former assumptions, gives 4.1 kev.

Still another way to look at the Thomas splitting is the following. We can think of the effect as being due not to the uniform precession of the P neutron, but rather, somewhat more realistically, to the short-lived, violent precessions accompanying two particle collisions. If we neglect the instantaneous motion of the center of gravity of the two particles, and take the force between them to be central, we obtain an interaction:

$$-\frac{\hbar^2}{16M^2c^2}\sum_{\text{pairs}}\left(\frac{1}{r^{ij}}\right)\frac{\partial V(r^{ij})}{\partial r^{ij}}(\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j)\cdot\mathbf{L}^{ij}, \quad (9)$$

where $\mathbf{r}^{ij}=\mathbf{r}^i-\mathbf{r}^j$ and $\hbar\mathbf{L}^{ij}=M[\mathbf{r}^{ij}\times(\mathbf{v}^i-\mathbf{v}^j)]$. The evaluation of this term is similar to others that will be discussed in the next section. The result has the same functional dependence and is of the same order of magnitude as (3) for a square well potential and a range 2.2×10^{-13} cm.

The radial integral in which $V(\rho)$ occurs is of the form $\int\rho^3d\rho(\partial V/\partial\rho)$. It is clear that the magnitude of the effect will depend greatly on the form of V. A $1/\rho^3$ dependence for $V(\rho)$ would lead to a divergent result; using this dependence, a cut-off could be adjusted to make the splitting as large as desired. However, potentials with leading term $1/\rho^3$ occur in mesotron theory only through the "tensor force." As will be shown in the next section the tensor force itself leads to a considerable splitting; with the adjustment of shape required here, it can be shown that it would lead to a much larger splitting than the Thomas precession, and in the opposite direction. It is therefore probably not possible to explain the magnitude of the doublet on the basis of relativistic spin-orbit precession, as customarily formulated.

† *Note added in proof.*—Stephens gives for the nucleus He^5-Li^5 a "Coulomb radius" of 4.2×10^{-13} cm (W. E. Stephens, Phys. Rev. **57**, 938 (1940)).

A possible modification of the calculation of the Thomas splitting involves simply multiplying the expression (9) by a constant. Breit[5] has obtained the Thomas force as a necessary part of the nuclear Hamiltonian by demanding covariance, to order v^2/c^2, of the Dirac wave equation for a nuclear particle. If the force field is taken to be a fixed external field, the Thomas term is uniquely fixed as $(\hbar/4c^2)\boldsymbol{\sigma}\cdot[\dot{\mathbf{v}}\times\mathbf{v}]$. However, when the forces are due to interactions between particles, there is an arbitrariness essentially in the choice of a multiplicative constant.

If the inter-particle forces are derivable from an "ordinary" potential $J(r^{ij})$, then the additive term necessary to guarantee covariance is

$$Q_0=-\frac{\hbar}{4Mc^2}\sum_{\text{pairs}}\{[\boldsymbol{\nabla}^i J(r^{ij})\times\boldsymbol{\sigma}^i]\cdot[\mathbf{v}^i b^{ij}+\mathbf{v}^i(1-b^{ij})]+[\boldsymbol{\nabla}^i J(r^{ij})\times\boldsymbol{\sigma}^i]\cdot[\mathbf{v}^i b^{ij}+\mathbf{v}^i(1-b^{ij})]\}, \quad (10)$$

where b^{ij} is, to this point, arbitrary. If we take $b^{ij}=1$ for all i, j, we get

$$Q_0=\frac{\hbar}{16Mc^2}\sum_{\text{pairs}}\{[\mathbf{v}_-{}^{ij}\times\boldsymbol{\nabla}_-{}^{ij}J(r^{ij})]\cdot[\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j]+[\mathbf{v}_+{}^{ij}\times\boldsymbol{\nabla}_-{}^{ij}J(r^{ij})]\cdot[\boldsymbol{\sigma}^i-\boldsymbol{\sigma}^j]\}. \quad (11)$$

Here $\mathbf{v}_+{}^{ij}=\mathbf{v}^i+\mathbf{v}^j$, $\boldsymbol{\nabla}_+{}^{ij}=\boldsymbol{\nabla}^i+\boldsymbol{\nabla}^j$, etc. This result is the intuitive extension of $(\hbar/4c^2)\boldsymbol{\sigma}\cdot[\dot{\mathbf{v}}\times\mathbf{v}]$ and agrees with (9) if the center of gravity is at rest: $\mathbf{v}_+{}^{ij}=0$. Q_0 is not essentially different if Majorana or Heisenberg forces are considered.

On the other hand, if we take $b^{ij}=b\gg 1$, we get

$$Q_0=\frac{b\hbar}{8Mc^2}\sum_{\text{pairs}}[\mathbf{v}_-{}^{ij}\times\boldsymbol{\nabla}_-{}^{ij}J(r^{ij})]\cdot[\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j]. \quad (12)$$

The velocity of the center of gravity is not involved. This is just the expression (9) multiplied by $2b$. b could be adjusted to give correctly the experimental value of the splitting; a value of the order of 10–100 would be needed. The weakness of such a method of treatment is that such a large value of b renders meaningless the expansion in terms of powers of v^2/c^2 whereby (10) was obtained. One can conclude only that large spin-orbit forces may be present, but not that they will necessarily be of the form (12).

[5] G. Breit, Phys. Rev. **51**, 248 (1937).

III

One would expect a contribution to this splitting from the "tensor force" of mesotron theory

$$T=\sum T^{ij}=\sum_{\text{pairs}}[(\boldsymbol{\sigma}^i\cdot\hat{\mathbf{r}}^{ij})(\boldsymbol{\sigma}^j\cdot\hat{\mathbf{r}}^{ij})-\tfrac{1}{3}\boldsymbol{\sigma}^i\cdot\boldsymbol{\sigma}^j]V(r^{ij}), \quad (13)$$

where $\mathbf{r}^{ij}$ is the vector joining the positions of two nuclear particles, $\boldsymbol{\sigma}^i$ and $\boldsymbol{\sigma}^j$ the corresponding spin vectors.[4] The radial function $V(r^{ij})$ occurs in mesotron theory as a singular expression with leading term $\sim 1/r^3$; to avoid divergences, it must either be cut off at short distances or replaced by a less singular function of depth and range empirically determined. The procedure here followed will be to use for $V(r^{ij})$ a square well of radius $a=e^2/mc^2$ and of depth 25 Mev, which combination results in the observed quadrupole moment for the deuteron.

As has already been pointed out, T gives no shift to either the $^2P_{1/2}$ or the $^2P_{3/2}$ level in first order, and hence gives no splitting between them.[6] This results simply from the fact that T, in its dependence on rotations involving either the spin or space coordinates separately, behaves as D_2, a representation of second rank of the rotation group. The He^5 ground state has spin $=\frac{1}{2}$; it can have no diagonal matrix elements for T.

We have now to consider the second-order contribution to the splitting. The intermediate states will be taken to be solutions of the unperturbed Hamiltonian, that is, they have definite S and L. They must have components in their reduction with $J=\frac{1}{2}$ or $\frac{3}{2}$ and must also have odd parity. Because of the selection rule on the tensor force, they must have spin $\geqslant\frac{3}{2}$. This can only be obtained if the α-particle core is broken up; consequently the combining states must all possess large excitations, probably >15 Mev.

The question arises as to whether one can obtain the second-order splitting by calculating explicitly the contribution of a relatively small number of upper states possessing, presumably, the least excitation. To test this possibility, we made a calculation of the depression contributed

[6] S. M. Dancoff, Phys. Rev. **56**, 384 (1939).

by a given upper state, namely that formed by raising one of the protons out of the α-particle to an excited S orbit. The resulting configuration has components which have spin $\frac{3}{2}$ as well as total angular momentum $\frac{1}{2}$ and $\frac{3}{2}$, and odd parity. If one takes 15 Mev as the excitation of this level, one finds that it contributes the following amounts to the depressions of the two ground states:

$$^2P_{1/2} - 23{,}000(a/\lambda)^{10} \text{ ev},$$
$$^2P_{3/2} - \ 1000(a/\lambda)^{10} \text{ ev}.$$

In this calculation, the radial wave functions were taken of the form $e^{-r/\lambda}$ for the S orbits and $re^{-r/\lambda}$ for the P orbit. From the considerations of II, one sees that λ should be of the order of $4-5\times10^{-13}$ cm; a is about 2.8×10^{-13} cm. Consequently, setting $a/\lambda=1$ should provide a generous upper limit for the above depressions. Not only is the splitting extremely small, but it is in the opposite direction to that experimentally observed.

The following procedure should now serve as a check on whether the inclusion of a few excited states such as the above is sufficient for a good estimate of the second-order effects. The depression of either level may be written:

$$\sum_n \frac{(u_a T v_n)(v_n T u_a)}{E_0 - E_n}, \tag{14}$$

where $(u_a T v_n)$ is the matrix element of T between the ground state wave function u_a ($a=\frac{1}{2}$ or $\frac{3}{2}$) and the excited state wave function, v_n. An upper limit to this quantity is given by

$$\frac{1}{E_0 - E_{\min}} \sum_n (u_a T v_n)(v_n T u_a),$$

where $E_{\min}$ is the energy of the lowest combining excited state. By closure, this expression becomes

$$(u_a T^2 u_a)/(E_0 - E_{\min}). \tag{15}$$

If it is true that only a few excited states are involved, this expression should also be quite small if a minimum energy denominator of ~15 Mev is used; it should be not more than a few times the 23,000 ev found above.

The quantity T^2 should now be expressed in terms of its irreducible components. Such a reduction will yield: first, terms which transform like D_0 of space and spin separately; second, terms transforming like D_1 of space and spin separately; and last, terms representing higher rank transformation properties, which can be ignored because they have no diagonal matrix elements for the states u_a. The following expressions are obtained:

For terms in T^2 of the form $T^{ij}T^{ij}$:

$$\begin{aligned} &D_0\text{: } \tfrac{2}{3}(1+\tfrac{1}{3}\boldsymbol{\sigma}^i\cdot\boldsymbol{\sigma}^j)\,V^2(r^{ij}), \\ &D_1\text{: vanishes.} \end{aligned} \tag{16}$$

For terms in T^2 of the form $T^{ij}T^{jk}$ $(i\neq k)$:

$$D_0\text{: } \tfrac{2}{3}\{\boldsymbol{\sigma}^i\cdot\boldsymbol{\sigma}^k\}\,\{(\mathfrak{l}^{ij}\cdot\mathfrak{l}^{jk})^2-\tfrac{1}{3}\}\,V(r^{ij})\,V(r^{jk}), \tag{17}$$

$$D_1\text{: } [\boldsymbol{\sigma}^i\times\boldsymbol{\sigma}^k]\cdot[\mathfrak{l}^{ij}\times\mathfrak{l}^{jk}](\mathfrak{l}^{ij}\cdot\mathfrak{l}^{jk})\,V(r^{ij})\,V(r^{jk}). \tag{18}$$

For terms in T^2 of the form $T^{ij}T^{kl}$ (no two indices alike): Terms of this sort can always be disregarded if one deals with a Hartree configuration in which one nuclear particle is in a P state, the other four in S states; one sees this by applying the selection rule for the tensor force to the integrals over the coordinates of the four particles.

An analysis of the integrals involved shows that the terms (17) and (18) have extremely small matrix elements for the ground states, giving a contribution to (15) of not over 10,000 ev for either state.[7] On the other hand, the terms

[7] The reason these terms are small is essentially the following. The transformation properties of $(\mathfrak{l}^{ij}\cdot\mathfrak{l}^{jk})(\mathfrak{l}^{ij}\times\mathfrak{l}^{jk})$ are $D_2(\mathfrak{l}^{ij})D_2(\mathfrak{l}^{jk})$. (The same is true of $(\mathfrak{l}^{ij}\cdot\mathfrak{l}^{jk})^2-\frac{1}{3}$.) This factor is to be integrated with a product of 3 initial and 3 final wave functions and with $V(r^{ij})\,V(r^{jk})$. We find that the product of 6 wave functions, if analyzed with regard to its transformation properties for rotations of $\mathfrak{l}^{ij}$ and $\mathfrak{l}^{jk}$, contains only a small amount of $D_2(\mathfrak{l}^{ij})D_2(\mathfrak{l}^{jk})$, resulting in approximate orthogonality.

Instead of the variables of integration $\mathbf{r}^i$, $\mathbf{r}^j$, $\mathbf{r}^k$, take $\mathbf{r}^{ij}$, $\mathbf{r}^{jk}$, $\mathbf{r}^j$. Functions of r^i (e.g.) that occur in the integral are expanded by means of the relation

$$r^i=[(r^j)^2+(r^{ij})^2-2r^jr^{ij}\cos\gamma]^{\frac{1}{2}} \sim r^j-r^{ij}\cos\gamma+(r^{ij})^2/2r^j-(r^{ij})^2\cos^2\gamma/2r^j+\cdots.$$

γ is the angle between $\mathbf{r}^j$ and $\mathbf{r}^{ij}$. Here we treat r^{ij} as smaller than r^j; this will be valid for most of the range of integration because of the fact that $V(r^{ij})$ vanishes for $r^{ij}>a=2.8\times10^{-13}$ cm, whereas the size of the wave functions is about $\lambda=4.5\times10^{-13}$ cm. If the wave functions are expanded for $r^{ij}<r^j$, the leading terms, namely those in $(\cos\gamma)^0$ and $\cos\gamma$ are orthogonal to $D_2(\mathfrak{l}^{ij})$; we must go to the terms in $\cos^2\gamma$, at the same time introducing a factor $(r^{ij})^2/(r^j)^2$ which results in an extra factor $(a/\lambda)^2$ in the result.

A similar expansion of the functions of r^k about $r^{jk}=0$ leads again to the elimination of two leading terms and to the introduction of a factor $(r^{jk})^2/(r^j)^2$.

Evaluating term (18) we find contributions on the order of $10^{-1}(a/\lambda)^{10}$ (Mev)^2. With an energy denominator of 15

(16) contribute in the neighborhood of 2 Mev to (15), this contribution being the same for both states. A comparison of this result with the upper limit of 23 kev found for the depression of the $\frac{1}{2}$ level due to a single excited state is indication that an extremely large number of states extending to high energies of excitation is important in the sum (14).

Equation (14) may always be rewritten, using an average energy denominator, in the form

$$(u_a T^2 u_a)/(E_0-\bar{E}_a). \tag{19}$$

As was shown above, the factor $(u_a T^2 u_a)$ can differ for the two states by not more than 10 kev, which is an upper limit for the contribution of the D_1 component of T^2. Therefore if the splitting of the two states by the tensor force is actually of the order of magnitude of the experimental result, it must be due almost entirely to the variation in $\bar{E}_a$ for the two states. In the following, to simplify the argument, we make the assumption that $(u_a T^2 u_a)$ is the same for $a=\frac{1}{2}, \frac{3}{2}$.

By writing the energy denominator in (14) in terms of a partial expansion about its mean value, one obtains the identity:

$$\frac{1}{E_0-E_n}\equiv\frac{1}{E_0-\bar{E}_a}+\frac{E_n-\bar{E}_a}{(E_0-\bar{E}_a)^2}+\frac{(E_n-\bar{E}_a)^2}{(E_0-\bar{E}_a)^2(E_0-E_n)}. \tag{20}$$

The contribution of the second and third terms to (14) vanishes, since this was the criterion for the choice of $\bar{E}_a$; the splitting of the two states is due to the variation of $\bar{E}_a$.

But now let us make the expansion (20) not about $\bar{E}_a$, but about $\bar{E}$, chosen *the same* for the two states. The first term in the expansion (20) now gives an equal contribution to (14) for the states u_a. The splitting has become buried in the second and third terms. We will calculate the splitting due to the second term, making the assumption that the third term may be made to contribute a vanishing splitting by an appropriate choice of $\bar{E}$. The possibility of such a procedure rests on the fact that the splitting due to the second term does not depend critically on the choice of $\bar{E}$; in particular it cannot change its sign as a function of $\bar{E}$. On the other hand, the splitting due to the third term does depend critically on this choice. It may be written

$$\int_{E_0+15}^{\infty} dE_n \frac{(E_n-\bar{E})^2}{(E_0-E_n)} G(E_n), \tag{21}$$

where the function $G(E_n)$ has zero average value. The factor $(E_n-\bar{E})^2/(E_0-E_n)$ is always negative and has a zero point at $E_n=\bar{E}$ which may in general be adjusted so as to reduce the value of the integral as much as desired. Our ignorance of the precise form of $G(E_n)$, that is to say of the spectra of $|(u_a T^2 v_n)|^2$, prevents the exact specification of this required value of $\bar{E}$. If the two spectra were very similar, but somewhat displaced, $\bar{E}$ would lie between $\bar{E}_{1/2}$ and $\bar{E}_{3/2}$. If one considers other forms which $G(E_n)$ might reasonably have, one finds values of $\bar{E}$ not much removed from the above. At any rate, the arguments that follow are essentially unaffected by the value of $\bar{E}$, within wide limits; it is only necessary that it can in principle be found.

One has now to calculate the splitting due to the expression

$$\sum_n \frac{(u_a T v_n)(v_n T u_a)}{(E_0-\bar{E})^2}(E_n-\bar{E}). \tag{22}$$

$\bar{E}$ may be dropped as not contributing to splitting. The remainder is

$$\sum_n \frac{(u_a T H_0 v_n)(v_n T u_a)}{(E_0-\bar{E})^2}=\frac{(u_a T H_0 T u_a)}{(E_0-\bar{E})^2}$$

by closure, where H_0 is the unperturbed Hamiltonian. Proceeding as before, we look for the part of TH_0T which transforms like D_1. We permit H_0 to include the kinetic energies plus a sum of potential terms which may represent ordinary forces, spin exchange, space exchange, and complete exchange forces. If we examine these potential terms we find that they do indeed lead to D_1 components for TH_0T, but only by combining $D_2(\hat{\imath}^{ij})$ with $D_2(\hat{\imath}^{jk})$, in other words forming structures of the form $(\hat{\imath}^{ij}\cdot\hat{\imath}^{jk})(\hat{\imath}^{ij}\times\hat{\imath}^{jk})$; it is immediately clear that no other space vectors could result from such combinations. The contribution of these terms is again very

Mev, this results in a depression (15) of less than $10^{-2}(a/\lambda)^{10}$ Mev. With the estimated values of the constants $a/\lambda=0.62$, and the result is reduced to less than 10^{-4} Mev. The upper limit of 10^{-2} Mev should be quite safe in this case.

small, for the same reason that the contribution of (18) is small.[7]

On the other hand, the kinetic energy terms in H_0 lead to space vectors of lower rank symmetry properties. The D_1 component of

$$T^{ij}\sum_{\mu}[-(\hbar^2/2M)\nabla_\mu{}^2]T^{ij}$$

is found to be

$$-\frac{i\hbar^2V^2(r^{ij})}{Mr^{ij}}(\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j)\cdot[\mathbf{r}^{ij}\times(\boldsymbol{\nabla}^i-\boldsymbol{\nabla}^j)], \quad (23)$$

which can be written

$$\frac{\hbar^2V^2(r^{ij})}{M(r^{ij})^2}(\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j)\cdot\mathbf{L}^{ij}, \quad (24)$$

where $\hbar\mathbf{L}^{ij}$ is the relative angular momentum of the particles i and j. The above "diagonal" terms in $T\sum_\mu[-(\hbar^2/2M)\nabla_\mu{}^2]T$ prove to be more important than the "nondiagonal" terms—those that involve different T^{ij}'s on left and right; the latter have D_1 components of higher symmetry and give a contribution of the order of the terms (18).

Consequently the leading terms in the splitting will be given by the difference in the matrix elements of

$$\frac{\hbar^2}{M(E_0-\bar{E})^2}\sum_{\text{pairs}}\frac{V^2(r^{ij})}{(r^{ij})^2}(\boldsymbol{\sigma}^i+\boldsymbol{\sigma}^j)\cdot\mathbf{L}^{ij}, \quad (25)$$

where $E_0-\bar{E}$ has a value which is not less than 15 Mev and probably not greater than 40 or 50 Mev; we will take 30 Mev as an estimate which should give the right order of magnitude.

The interaction (25), in contrast to the Thomas interaction, (9), is a "normal" spin-orbit force; that is to say, one would expect it to give rise to a splitting in which the state of higher multiplicity lies higher, because for such a state the spin of a particular particle and its angular momentum relative to any other particle would have the maximum tendency to be aligned. The calculation verifies this conclusion and gives for the splitting $80(a/\lambda)^5$ kev, where as before a is the range of the tensor force and λ the scale of the nuclear wave functions; the state $J=\frac{1}{2}$ is found to lie lower.

It is clear from the form of (25) that the magnitude of the result will be extremely sensitive to the form of $V(r^{ij})$. The radial integral in which V occurs is of the form $\int\rho^2d\rho V^2(\rho)$ so that the $1/\rho^3$ dependence of mesotron theory would give an extremely bad $1/\rho^4$ divergence. If one uses instead of the square well potential the cut off potential proposed by Bethe,[8] and fitted by him to the two levels of the deuteron, one obtains a much larger result. Using the neutral mesotron theory with zero cut-off inside the radius $x_0=0.318a$, one obtains $800(a/\lambda)^5$ kev for the splitting of the two levels, the order being of course as before. It is clear, then, that splitting of the order of magnitude of the experimental result could be caused by the tensor force.

The terms which were discarded in comparison with (25), and whose upper limit was said to be 10 kev, (e.g. (17), (18)) prove to be much less affected by the shape of V. The radial integrals of interest are of the form $\int\rho^4d\rho V(\rho)$; they would not diverge if a $1/\rho^3$ potential was inserted. If Bethe's cut off potential is used, the upper limit previously assigned is practically unaffected.

As a result of this section it is indicated that the tensor force gives rise to a splitting in the opposite direction to that experimentally observed and of a magnitude which is extremely sensitive to the shape of the force but which is probably of the order of 100 kev. The direction of the splitting is of course independent of any ambiguity in the sign of the tensor force in the Hamiltonian, since in second order it occurs quadratically. The importance of higher order contributions is difficult to evaluate since it will depend to a great extent on the degree of "peakedness" assigned to the potential function. Because of this, as well as the considerable experimental uncertainty, the conclusion which may be drawn must be regarded as tentative. It is that the structure of the He^5 doublet is not determined by the tensor force, and can be assigned to Thomas relativistic coupling only with an adjustment of multiplicative constant such as described in II.*

The author is indebted to Professor J. R. Oppenheimer for contributing many of the ideas presented herein. The helpful interest of Dr. L. I. Schiff is also gratefully acknowledged.

[8] H. A. Bethe, Phys. Rev. **57**, 390 (1940).

* See "Note Added in Proof" above.

Nuclear Isotope Shift in the Spectra of HgH^+ and HgD^+

S. MROZOWSKI*
Institute of Theoretical Physics, Joseph Pilsudski University, Warsaw, Poland
(Received June 10, 1940)

The structure of the lines in the band spectra of HgH^+ and HgD^+, excited in a hollow cathode tube, is studied with Fabry-Perot etalons. The lines are found to be broadened, with traces of structure, an effect caused by the isotopic constitution of mercury. The structure shows that there is no anomalous behavior of components belonging to the odd Hg isotopes due to hyperfine splitting, but the half-widths of the lines differ widely from the expected ones. From the measured half-widths the shifts between the consecutive even isotopes of Hg are evaluated and then the differences between these and the values obtained from the formulas for the normal vibrational and rotational isotope effect calculated. These additional shifts show the same behavior as the shifts found some years ago in HgH and HgD and therefore must be explained as nuclear isotope shifts, that is, shifts connected with differences in nuclear volume of the different isotopes of Hg. For the lowest vibrational bands the shifts are of the same sign as in HgH, but decrease with the vibrational energy of both upper and lower states, changing sign for sufficiently high vibrational energies and probably tending in the limit of dissociation to the shifts found in the corresponding quadrupole line 2815A of Hg II. Consideration of the data on the nuclear isotope shifts in HgH, HgH^+, Hg I and Hg II permits some conclusions about the deformation of the electron clouds in Hg and Hg^+ caused by the attachment of a hydrogen atom, and a rough calculation of the nuclear shifts in the molecular levels of HgH and HgH^+.

THAT different isotopes of the same element emit spectra in which the lines are relatively shifted is well known in atomic spectra. The shifts found in the light elements are caused by differences in the masses of the nuclei; in proceeding from the lighter to heavier elements these mass isotope shifts are found to become rapidly smaller, and for Mg are difficult to observe. For elements heavier than argon the shifts are so small that they can be of no importance experimentally and therefore the shifts found in the spectra of heavier elements must be explained in a different way. Probably these latter shifts are caused by differences in the distribution of the electric field in the immediate neighborhood of the nuclei, i.e., in the region where the field differs considerably from the Coulomb field. Following a suggestion of Professor Oppenheimer, I shall hereafter call these *nuclear* isotope shifts. For several years I have been interested in the question as to whether the influence of other nuclear properties besides mass could be detected in band spectra. For example, one would expect differences in the electric fields to influence the positions of lines, and the presence of magnetic moments to produce hyperfine structure. Both problems are more complicated than in atomic spectra. In the former case, with which we shall be mainly concerned in this paper, the chief difficulty is that the mass shifts of the corresponding lines of isotopic molecules are considerable, and even in the most favorable cases cannot be neglected. But a similarity with the case of atomic spectra is revealed if we consider the *differences* between the observed shifts and the ones calculated by the well-known formulas for the rotational and vibrational isotope effects. For isotopic molecules differing very much in reduced mass (i.e., when a light atom in the molecule is exchanged for its isotope) the true mass shift differs from the shift given by the above-mentioned formulas. Differences in the rotational and vibrational effects, as well as shifts of whole band systems, have been found in many spectra, especially in the hydrides and deuterides, during the last eight years. As was shown by Van Vleck[1] and others, their magnitudes are in agreement with calculations based on general quantum-mechanical theory. These departures from the usual formulas diminish with the differences of reduced mass so rapidly that even for the isotopic pair $C^{13}N^{14}$ and $C^{12}N^{14}$ excellent agreement with the usual formulas has been obtained.[2] But some years ago additional

* Now at Radiation Laboratory, Department of Physics, University of California.

[1] J. H. Van Vleck, J. Chem. Phys. **4**, 327 (1936).
[2] F. A. Jenkins and D. E. Wooldridge, Phys. Rev. **53**, 137 (1938).

shifts were discovered by the author[3] in the case of isotopic molecules $Hg^{198}H$, $Hg^{200}H$, $Hg^{202}H$, and $Hg^{204}H$, for which the mass differences cannot be responsible, the shifts found being more than a hundred times larger than the expected mass correction shifts. These shifts have been explained as nuclear isotope shifts. Bohr[3] has shown how the correlation with the shifts in the atomic spectrum of Hg as well as the observed dependence of this effect on the vibrational and possibly also rotational energy of the molecule is to be understood. A confirmation of this effect has been found in the case of the HgD spectrum,[4] but up to the present it has been detected in no other spectrum. It was pointed out[4] that if the explanation is correct, the investigation of the spectra of the molecules HgH^+ and HgD^+ would be very promising for this purpose. Consequently experiments were performed with a Lummer-Gehrcke plate for high resolving power, but in view of the low intensity of this spectrum in the mercury arc I have not been able to photograph it in a reasonable exposure time. Recently, while performing some experiments with the spectrum of boron[5] I succeeded in examining the pattern of some exceedingly weak lines in the region of 2400A arising from Fe contamination. This was made possible by the relatively great transparency of Hochheim amalgam used on the Fabry-Perot plates. The etalon also proved to have sufficient resolving power in the farther ultraviolet, as the study of the fivefold structure of the line 2262A of HgII has shown. On the other hand, it was shown by Almy and Horsfall[6] that the spectrum of HgH can be excited in a hollow cathode discharge with great intensity, and therefore a new attempt to analyze the structure of the lines of the spectra HgH^+ and HgD^+ was made. This time the experiments were successful and the results are reported here.

The band spectra of HgH^+ and HgD^+ were excited in a hollow cathode discharge tube, with a d.c. power supply described in detail in an earlier paper.[5] The tube has proved to be very efficient for producing intense spectra, as experiments with B,[5] Be,[7] HgII,[8] HgH^+ and HgD^+, and ZnH[9] have shown, and therefore it seems worth while to give a more detailed description of this tube, in spite of the fact that during recent years many hollow cathode discharge tubes of different shapes have been reported. The tube is especially suitable for experiments in which high electric currents must be used. It has been tried with 2.5 amp. for short intervals of time and with continuous running at 1.5 amp. A schematic diagram

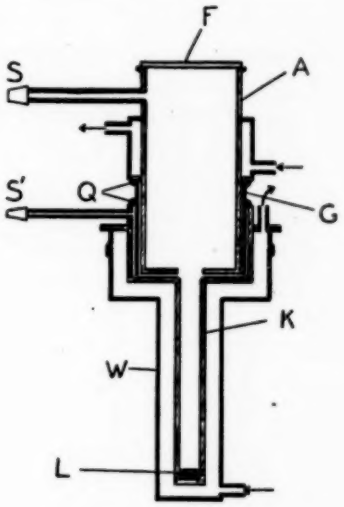


FIG. 1. Diagram of discharge tube.

of the tube is given in Fig. 1. The water-cooled anode *A* and the removable water jacket *W* are of brass and the cathode *K* (about 7 cm long and 8 mm in diameter) of pure copper. As has been previously pointed out,[5] copper cathodes are especially advantageous in view of the purifying action of the clean-up. The anode is separated from the cathode by a gap of about 1 mm and they are held in position by a glass cylinder *G*. The smallness of the gap is very important when high currents are used, because in case of larger separations the discharge changes suddenly into an arc burning along the walls of the glass cylinder, which after a short while causes the glass to crack.[10] The glass cylinder is sealed to the anode and cathode with Apiezon Q wax. With

[3] S. Mrozowski, Zeits. f. Physik **95**, 524 (1935).
[4] S. Mrozowski, Zeits. f. Physik **99**, 236 (1936).
[5] S. Mrozowski, Zeits. f. Physik **112**, 223 (1939).
[6] G. M. Almy and R. B. Horsfall, Phys. Rev. **51**, 491 (1937).
[7] S. Mrozowski, Phys. Rev. **55**, 798 (1939).
[8] S. Mrozowski, Phys. Rev. **57**, 207 (1940).
[9] The results of these experiments will be reported in the near future.
[10] Cf. R. M. Elliott and J. Wulff, Phys. Rev. **55**, 170 (1939). These investigators solved the difficulty in the opposite way—that is, by enlarging the gap. Their solution does not give the high electric field desired in the neighborhood of the hollow part of the cathode.

very high currents (more than 1.5 amp.) this wax softens after a short while and in case of steady running, to prevent the heating of the glass cylinder, the anode should probably be cooled from the inside with water down to the bottom. For liquid-air cooling the lower wax joint must be heated with running water. A special water circuit for this purpose was arranged in one of the cathodes used. The discharge tube was connected with the pumping system at S and with the gas supply at S'. From S' the gas was let into the tube through a groove made in the cathode wall and extending beyond the bottom of the glass cylinder. If gas circulation is needed this groove can be extended along the walls to the bottom of the hollow part of the cathode and covered with long plates tightly fitting so as to allow the gas to flow out only at the bottom of the cathode. This device was used only at the start of these experiments, when an iron cathode was used because the purifying effect of the clean-up on copper was not then appreciated.

A large drop of mercury was placed in an iron cup at the bottom of the cathode and during the experiments the mercury slowly vaporizes, producing a layer of amalgam on the walls of the hollow cathode, but most of it condenses at the bottom of the anode. The latter gives some trouble, because when the drops of mercury grow too large they short-circuit the discharge. The tube must then be opened and the anode cleaned by wiping off the liquid mercury. After setting the tube back into position the discharge runs unsteadily for some time, changing frequently into an arc in the small gap, until all impurities are removed by the discharge from the surface of the cathode.

The discharge occurred in a mixture of helium and hydrogen (or deuterium). Spectroscopically pure helium was admitted in small quantities by a suitable arrangement of stopcocks and as a source of hydrogen a platinum-palladium tube heated by a gas flame was used. Deuterium was obtained from 99.2 percent heavy water by electrolysis and then purified by diffusing it through palladium. A very useful arrangement for this purpose, which is a modification of one used by Professor W. Dziewulski in Wilno, is shown in Fig. 2. It makes possible the admission of any desired small quantity of the purest deuterium

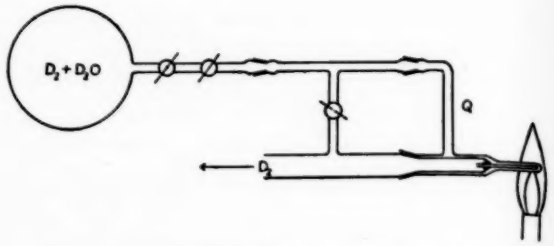


FIG. 2. Arrangement for admitting gas. Q is a fused silica cap.

while by removing the fused silica cap Q it can also be used for the introduction of hydrogen. In most of the experiments helium at about 1 mm pressure mixed with H_2 (or D_2) at a pressure of some tenths of a mm was used. This was found to be the best condition for obtaining an intense HgH^+ spectrum. During operation the hydrogen disappeared by absorption (chiefly in the mercury deposited on the anode), and the proper pressure in the discharge tube was maintained by diffusion from a large container connected with the tube and filled with the $He+H_2$ mixture. But the absorption was so rapid that a new portion of hydrogen had to be added every 10–15 minutes. In view of the slowness of the diffusion along the thin tubes connecting the container with the discharge tube, the hydrogen pressure in the hollow cathode was certainly much lower than in the container.

The discharge showed some interesting features. Three types of discharge were observed. When a potential difference is applied which is insufficient for starting the usual glow discharge, only a very weak luminosity in the center of the cathode can be seen. The current is undetectable on an instrument reading to 1/100 amp. This type of discharge is of no use for spectroscopic purposes. If we increase the voltage the usual glow discharge starts suddenly, and if we immediately lower the voltage to such an extent that the current is not more than about 0.3 amp. the glow discharge runs steadily. With higher currents (the limit varies according to circumstances from 0.25 to 0.35 amp. and more) the discharge is unsteady, and after a short while suddenly changes into a new type, apparently not previously described. A sudden increase of current (by about 2 to $2\frac{1}{2}$ times) is accompanied by a corresponding increase of intensity of the cathode

glow. This new type has some of the features of an arc discharge, because the mercury drop evaporates very quickly from the iron cup (in about 15 minutes) and the discharge emits a quite different spectrum. No helium lines occur at all, while the Hg, H_2 and HgH lines have very high intensity. But that it is not an ordinary arc is shown by the relatively low temperature. This is not appreciably higher than in the usual glow discharge as can be seen from the distribution of intensities in bands of HgH and HgH^+. In an ordinary mercury arc running on low current the temperature is much higher, since many more lines are observed in branches and higher vibration bands are emitted. Very often this new type of discharge changes suddenly into an arc discharge burning between the upper amalgamated end of the cathode and the anode. In this case a cathode spot is formed, moving rapidly around the cathode, and sometimes going into the hollow part, but never deeper than about 1 cm. This occurs with higher currents while for lower currents the discharge runs with frequent changes back into the ordinary glow discharge.

The condition of the inner surface of the cathode has a very strong influence on the above phenomena. The effects described were observed for cathodes which had been in long use and which had a heavily amalgamated inner surface. The discharge ran in helium with an admixture of hydrogen or in pure helium. If the surface does not contain mercury, as in case of a Fe cathode or a fresh Cu cathode, the limits of instability are much higher, say above 1 amp. The spectrum of the high current type does not differ from that described above, but in the glow discharge the intensity of the HgH bands is very low. The behavior of the bands HgH^+ was not studied, but probably is similar. The intensity distribution is normal in the case of a fresh Cu cathode, but in the case of a Fe cathode shows an abnormally low temperature distribution, only a few lines in each branch being observed. Such abnormal distributions were found by Schüler and his co-workers[11] in other hydride spectra from the hollow cathode. They gave an explanation based on a mechanism of the formation of molecules at the surface of the cathode. The results reported here undoubtedly show the importance of the state of the surface, but seem to indicate that the relations are more complicated than assumed in their explanation.

The spectrograms of the HgH^+ and HgD^+ bands were mostly obtained with the ordinary glow discharge. The new type of discharge was not used for several reasons: The greater steadiness of the glow discharge; the necessity of frequent refilling with fresh mercury; the slightly higher temperature and consequent lack of sharpness of the lines (in the hollow cathode about 20 lines and in the new discharge type one or two more in both branches of the (0,0) band of HgH^+ were observed, in contrast to the usual low current mercury arc, where more than 30 lines have been reported[12]); and finally perhaps the most important reason was the very high intensity of the continuous spectrum of hydrogen in the new type of discharge. This continuum is much more intense in the case of deuterium,[4] and the investigations are still more difficult. The continuous spectrum is especially unfavorable with the Fabry-Perot etalon used which gives, in contrast to the Lummer-Gehrcke plates, a strong continuous background for every line of about 8 percent of the maximum intensity of the line.[5] The Fabry-Perot etalon was used with fused silica spacers of from 7 mm to 30.2 mm. The patterns were photographed in a large interchangeable Hilger spectrograph having a high luminosity quartz optical system. Ilford Monarch and Wellington XPress photographic plates were used. The time of exposure of $\frac{1}{2}$ to $1\frac{1}{2}$ hours was within reasonable limits, so that no special arrangement for constancy of temperature was needed.

On the first exposures no structure could be found in the lines, all lines showing excessive broadening, especially those lying in the longer wave-length region of the spectrum. The lines of Hg II observed under the same conditions and in the same spectral region showed smaller widths which were in agreement with the calculated

[11] H. Schüler and H. Haber, Zeits. f. Physik **112**, 614 (1939) and other references there given.

[12] The band spectrum of HgH^+ was first found by H. Jeżewski, J. de phys. et rad. **9**, 278 (1928) and later analyzed more correctly by T. Hori, Zeits. f. Physik **61**, 481 (1930). The spectrum of HgD^+ was found and analyzed by the author, Acta. Phys. Polonica **4**, 405 (1936). The analysis of Hori has been corrected and improved constants for both HgH^+ and HgD^+ have been given in a paper by S. Mrozowski and M. Szulc, Acta Phys. Pol. **6**, 44 (1937).

Doppler widths. The latter were obtained by using the temperature of the discharge estimated earlier from the broadening of the boron lines.[5] Therefore the excessive broadening of the HgH^+ and HgD^+ lines was clearly due to some other cause, probably to the isotope effect. A more careful examination of the lines in the (0,0) band revealed that the width of the lines was greatest in the neighborhood of the band head and decreased in the direction of longer wave-lengths, a behavior directly opposite to the expected influence of the normal isotope effect (combined vibrational and rotational effect). This observation led me to measure as carefully as possible the half-widths of the lines. This was done by making series of photographs with a stepwise increase of the plate separation and measuring the breadths of the fringes with a comparator. No photographic photometry was applied because of the time required. But my experience with photometric measurements of the intensity distribution in the boron lines[5] showed the reliability of such direct estimates. The values found for different plate separations differed by no more than 10 percent and the mean values probably differ from the true half-widths by still less. The relative widths are naturally more exact. In the (1,2) band the results have been checked by finding the smallest plate separation for which the interference pattern is continuous, in which case the separation of consecutive orders is equal to the half-width of the line. This estimate is not as reliable as the former ones because the traces of structure found in the wider lines (see below) make the exact point of merging difficult to determine. Some results of these measurements are shown in Table I. In order to get some information from these values about the isotope effect, curves of the intensity distribution were drawn as in Fig. 3. A Doppler shape and a half-width 0.05 cm^{-1} were assumed. The lines due to different isotopes of mercury were taken to have isotope shifts of equal magnitude between the consecutive isotopes. This last assumption holds strictly for the normal isotope effect (vibrational and rotational), but for nuclear isotope shifts the components due to the odd isotopes lie closer to the components due to the even isotopes one unit lighter. This dissymmetry influences the shape of a line considerably when the nuclear isotope shift is greater than the normal one, as happens in a great part of HgH and HgD bands formerly studied. From the calculations given below it follows that in the present case the broadening of the lines is principally due to the normal effect (excluding the case of (0,0) bands, there the dissymmetry again plays no role, but this time because of the relatively larger Doppler broadening) and the components of the odd isotopes lie nearly midway between those of the even isotopes, as was assumed in Fig. 3.

As was previously pointed out[3] the dissymmetry of the nuclear isotope shift is not sufficient to explain the surprising absence of the odd components in the bands of HgH and HgD. A provisional explanation was given in assuming hyperfine structure splitting of these components. That their apparent absence is not at all connected with the nuclear isotope shift follows from the investigations of Svensson[13] in the spectrum of CdH, where the components of the odd Cd isotopes are again missing and where the nuclear isotope shift is negligible in comparison to the normal shifts. The supposition of a hyperfine structure splitting found a strong support in the presence of additional components observed in $^2\Sigma^* \rightarrow {}^2\Sigma$ bands of HgH[3]. Possibly the additional

[13] E. Svensson, *Die Bandenspektren des Cadmiumhydrides und -deuterides* (Ivar Hæggström Boktrykeri A. B., Stockholm, 1935), p. 91.

TABLE I. *Width of lines in* HgH^+ *and* HgD^+.

HgH^+				HgD^+			
BAND	LINE	MEASURED WIDTH	CALC. NORMAL EFFECT WIDTH	BAND	LINE	MEASURED WIDTH	CALC. NORMAL EFFECT WIDTH
(0,0)	44,058 cm^{-1}	0.077 cm^{-1}	0.058 cm^{-1}	(0,0)	44,060 cm^{-1}	0.079 cm^{-1}	0.065 cm^{-1}
(0,0)	43,712	0.059	0.075				
(0,1)	42,110	0.085	0.135	(0,1)	42,670	0.125	0.19
(0,1)	41,916	0.115	0.153				
(1,2)	41,723	0.17	0.155	(1,2)	42,400	0.175	0.22
(1,2)	41,619	0.18	0.168				

components found in CdH could be explained in the same way although the whole question seems to be more uncertain in this case in view of the possibility pointed out by Svensson of the existence of Cd isotopes not yet detected by the mass spectrograph. The point has been made,[4] that if the apparent absence is connected with hyperfine structure, it should not appear in bands of the type $^1\Sigma \rightarrow {}^1\Sigma$. This agrees with the results on HgH^+ and HgD^+. If the components of the isotopes 199 and 201 should be missing or at least if their intensity were distributed over a larger region, the minimum in the last curve in Fig. 3 would be more pronounced and a splitting of the lines of the (1,2) band into two groups easily observed. The experiments gave only slight traces of structure, corresponding quite exactly to the curves of Fig. 3, and giving support to the correctness of this explanation. Quite recently conclusive evidence for the existence of hyperfine structure in band spectra has been obtained in a reinvestigation of the HgH bands under much improved conditions. New components and differences in structure of the *R* and *Q* lines have been found. These results will be published as soon as microphotometer records of the plates have been evaluated.

On the basis of the curves of Fig. 3 the total isotope shift between the even isotopes of Hg (for instance, 202 relative to 200) can be found by interpolation in the case of the lines whose total half-widths are given in Table I. The values found are presented in Fig. 4. Their accuracy is probably greater than 10 percent and, as can easily be seen from the curves of Fig. 3, is practically independent of the assumption made about the symmetrical position of the odd isotopes. The continuous lines in Fig. 4 give the normal isotope shift (vibrational and rotational effect) calculated from the constants previously found by the author in collaboration with M. Szulc.[12] That the difference of the measured and calculated shifts given in Fig. 4 by thick arrows constitutes a real effect can be seen from Table I, where the widths of the lines calculated for the normal isotope effect are included for comparison. The differences are in some cases so far outside the limits of experimental error that there can be no question about the presence of a nuclear isotope shift. The accuracy of the measured half-

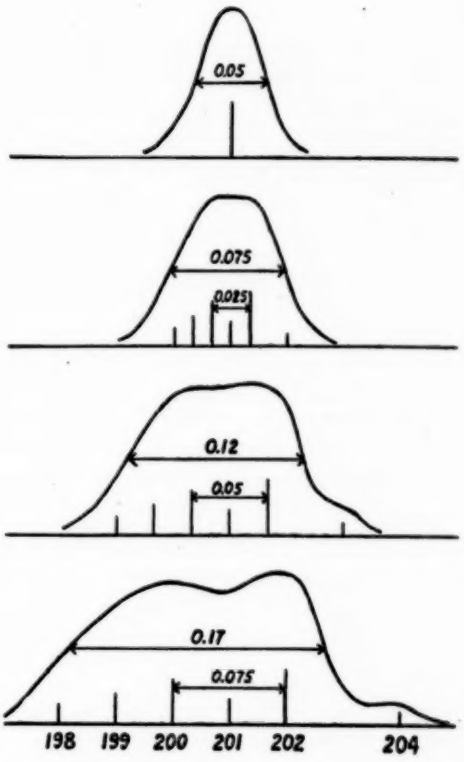


FIG. 3. Intensity distribution in spectral lines.

widths is shown by the equality of the inclination of the straight lines in Fig. 4, which shows the observed and calculated dependence of the isotope shift on the frequency of the lines in each band of HgH^+.

The differences between the observed and calculated values (Fig. 4) represent the additional isotope shift of the band origins, which are of the same order of magnitude as in HgH and HgD, though considerably smaller. As in the case of HgH they are some orders of magnitude greater than the mass correction shifts of the system origins, especially in the case of a $^1\Sigma \rightarrow {}^1\Sigma$ transition. It was shown[4] that in HgH and HgD these shifts are linear functions of the vibrational energy in both lower and upper states. Their constancy in the $^2\Pi$ state does not contradict this. Furthermore, it was found that this dependence is approximately the same for the hydride and deuteride, although differences were found lying somewhat outside the experimental error. Unfortunately in the HgH^+ and HgD^+ spectra only three bands of each molecule were observed, and the applicability of these rules cannot be proved

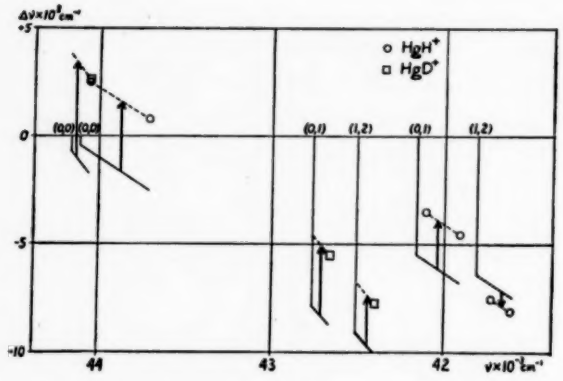


FIG. 4. Shifts of spectral lines. Differences between measured and calculated shifts are shown by the arrows.

directly. Nevertheless the three values allow one to calculate the three coefficients in the linear equation $\Delta\nu = a - b_1E_1 + b_2E_2$, and the approximate equality of the coefficients for the two spectra can be considered as an indirect proof of the correctness of the formula. The equations found are

$$\Delta\nu = 0.0450\ \text{cm}^{-1} - 12\times10^{-6}E_1 - 5.6\times10^{-6}E_2 \quad (HgH^+)$$

$$\Delta\nu = 0.0503\ \text{cm}^{-1} - 7.6\times10^{-6}E_1 - 5.7\times10^{-6}E_2 \quad (HgD^+)$$

where $\Delta\nu$ gives the shift of the component of the isotope 200 relative to 202 and the vibrational energy E is expressed in cm^{-1}. The difference in the coefficients b_1 is probably due to the fact that the shift for the (1,2) band of HgD^+ is in error. In general the measurements for HgD^+ are less reliable than for HgH^+ and in the case of the (1,2) band of HgD^+ there are so many lines close together that the value of the nuclear shift is very uncertain.

The shifts decrease for both upper and lower states. If they are true nuclear isotope shifts the value of $\Delta\nu$ must tend in the limit of dissociation of both upper and lower states to the value corresponding to the nuclear isotope shift in the corresponding atomic line. As was shown by Hori[12] and subsequently confirmed by the author[12] the atomic line in question is the forbidden quadrupole line 2815A of HgII[14] recently studied by the author.[8] The nuclear isotope shift amounts to about $-0.28\ \text{cm}^{-1}$ for 200 relative to 202. Instead of this we obtain by extrapolation of the linear formulas for HgH^+ and HgD^+ about -0.153 and $-0.120\ \text{cm}^{-1}$, respectively, if we assume as dissociation energies 1.0 volt for the upper state, $^1\Sigma^*$, and 2.15 volts for the lower state, $^1\Sigma$. The latter values were obtained from the constants of the band spectra[12] and were corrected to agree with the frequency of the corresponding atomic line, the correction being equally distributed between the two values. The difference between the shift found for the atomic line and the extrapolated band spectrum value is not surprising in view of the very small accuracy of such long extrapolations. It shows without question that in the neighborhood of the dissociation limit the linear rule breaks down. In the lower $^1\Sigma$ state this is probably connected with the detachment of one of the $6s\sigma$ electrons, which goes over to the hydrogen atom removed, just as occurs in the upper state $^2\Pi$ of HgH. In HgH the extrapolated value[4] was about one-half of the shift observed in the corresponding atomic line 2537A of HgI, very much as in the case here reported. We can therefore assume that here also the whole difference is caused by the detachment of the $6s\sigma$ electrons, an assumption which seems to be in agreement with the well-known strong dependence of the nuclear isotope shift on the presence of s electrons. The resulting nuclear shifts $\delta\nu$ of the minima of the potential curves are collected, together with other data for the molecules HgH^+ and HgH, in Table II. In this table the values given for the shifts of the atomic levels of HgII and HgI are based on the assumption of zero shift for the configuration $(5d)^{10}$. As is well known, the nuclear shifts are proportional to the densities of the electron cloud in the neighborhood of the nucleus. The data recently obtained from the spectrum of HgII[8] show the great influence of holes in the $(5d)^{10}$ shell on the electron density in the vicinity of the nucleus (see Table III, where the value $\sim0.285\ \text{cm}^{-1}$ was obtained from some as yet unpublished and very rough measurements of the lines 2916A and 5162A in HgII), the presence of electrons in levels higher than $6s$ playing a comparatively small role. From the

[14] Some predictions made in the paper (reference 4) were incorrect, being based on the correlation given in Mulliken's article in Rev. Mod. Phys. **4**, 80 (1932). The true correlation of the states is similar to that given by Mulliken for AgH and AuH in another paper, Phys. Rev. **33**, 735 (1929).

values of Table III it follows that the contribution of the first electron in an s shell is considerably greater than of the second one. In discussing the level shift in HgH and HgD in a former paper[4] equal contributions of 0.178 cm^{-1} for both s electrons were assumed because of insufficient data, and on this basis the lowering of the density of all electrons in the neighborhood of the Hg nucleus caused by the presence of a proton at a distance of 1.59A was calculated to be 26 percent of the density of both $6s$ electrons. With the exact values now available this now becomes 23 percent. Another assumption was made in this discussion concerning the nondeformation of the shell of $5d$ electrons. This is a very poor approximation because the proton has an equilibrium position deep inside of the electron cloud surrounding the Hg atom. That this assumption is not in agreement with experiment can be seen from Table II. The configurations outside the $5d$ shell show some similarity in the two pairs of states $^1\Sigma^*$ and $^2\Sigma$, $^1\Sigma$ and $^2\Pi$; hence in case of nondeformability of the $5d$ shell a negative nuclear isotope shift in the spectra of HgH^+ and HgD^+ would be expected. The observed positive effect arises from several causes: In the $^1\Sigma^*$ state relative to the $^2\Sigma$ state the lack of a $5d$ electron strengthens the binding of the $6s\sigma$ electrons thus increasing the level shift, but at the same time the smaller distance of the proton and the absence of the density of one $5d\sigma$ electron work in the opposite direction. There probably results a diminution of the level shift, because the value 0.410 found by extrapolation for $^1\Sigma^*$ appears to be too high from comparison of the other two states. That is, it seems quite improbable that the absence of a $6p\pi$ electron could strengthen the binding of the $6s\sigma$ electron enough to give an increase of the shift by 0.144 cm^{-1}—an increase by not more than 0.050 cm^{-1} seems reasonable. Therefore both values for the HgH^+ states are to be lowered by at least 0.100 cm^{-1}. The resulting approximate values are given in the last column of Table II. That they are quite reasonable can be seen by comparison of two different pairs of states. In $^1\Sigma$ the internuclear distance is smaller than in $^2\Sigma$. The distortion of the electronic cloud is larger but the absence of one electron increases the density to nearly compensate for this. In the $^1\Sigma^*$ state the r_e is larger, and the stronger binding of the $6s\sigma$ electron (than in $^2\Pi$) is nearly compensated by the absence of the density due to a $5d\sigma$ electron. From these considerations the somewhat unexpected fact follows that the density of the $5d\sigma$ electron must be relatively quite large.

The accuracy of the linear extrapolation is very low, except for the $^2\Sigma$ state of HgH where the measurements extend nearly to the limit of dissociation. The values of the coefficients b must be considered only as rough approximations. Nevertheless a comparison of the values given in Table II shows the dependence on the dissociation energy of the state previously pointed out,[4] namely for larger dissociation energies the absolute value of b is smaller. For an ionized molecule the b values are larger than for a neutral one, as was expected. All in all the relations in the nuclear isotope shift in the spectrum of HgH^+

TABLE III. *Contributions to shifts.*

CONFIGURATION	$5d^{10}6s^2$	$5d^96s^2$	$5d^86s^2$
Total shift	0.404 cm^{-1}	0.508 cm^{-1}	~0.750 cm^{-1}
Contribution of the 1th electron	0.226	~0.285	—
Contribution of the 2nd electron	0.178	~0.223	—

TABLE II. *Comparison of shifts of lines.*

MOLECULAR STATE	r_e A	D_{DISS} VOLT	$b\cdot10^6$	CORRESPONDING ATOMIC STATE	$\delta\nu_{ATOMIC}$ cm^{-1}	$\delta\nu_{MOL}$ cm^{-1}	(LIMIT) cm^{-1}	$\delta\nu_{M.\ CORR.}$ cm^{-1}
HgH^+								
$5d\sigma5d\pi^45d\delta^44s\sigma^26p\sigma\,^1\Sigma^*$	1.69	1	+12	$5d^96s^2\,^2D_{5/2}$	0.508	0.410	(0.508)	≦0.310
$(5d)^{10}6s\sigma^2\,^1\Sigma$	1.59	2.15	−5.6	$5d^{10}6s^2S_{1/2}$	0.226	0.455	(0.355)	≦0.355
HgH								
$(5d)^{10}6s\sigma^26p\pi^2\Pi$	1.59	1.94	~0	$5d^{10}6s6p\,^3P$	0.226	0.311	(0.311)	0.311
$(5d)^{10}6s\sigma^26p\sigma^2\Sigma$	1.73	0.37	+8	$5d^{10}6s^2\,^1S_0$	0.404	0.372	(0.404)	0.372

can be interpreted, although they are not as clear as in the case of HgH.

My best thanks are due to Professor C. Bialobrzeski for his continued interest in the progress of this work. This paper was submitted in August of last year for publication in the *Acta Physica Polonica*. Subsequent events have made it impossible to proceed with the printing of this journal, and the paper has therefore been rewritten and is presented here. I am very much indebted to Professor F. A. Jenkins for the kind help he has given me in the writing of this paper.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

The Ionization and Dissociation of Water Vapor and Ammonia by Electron Impact

MARVIN M. MANN, ANDREW HUSTRULID AND JOHN T. TATE
University of Minnesota, Minneapolis, Minnesota

(Received May 9, 1940)

The results of a mass-spectrometric study of the products of ionization and dissociation of water vapor and ammonia by electron impact are given. The ionizing potential of the H_2O molecule is found to be 13.0 ± 0.2 volts, and of the NH_3 molecule 10.5 ± 0.1 volts. The possible processes responsible for the formation of the various ions observed, both positive and negative, are discussed. Ions observed in H_2O vapor are H_2O^+, OH^+, O^+, H^+, H_2^+, H_3O^+, O^-, and H^-. Ions observed in NH_3 are NH_3^+, NH_2^+, NH^+, N^+, H^+, NH_3^{++}, NH_2^-, and H^-.

INTRODUCTION

THE use of the mass spectrometer for the study of products of ionization and dissociation has come to be recognized as a valuable method of attack on the problem of molecular structure.[1-4] In view of the excellent and thorough exposition of the method in previous papers, particularly that of Smith,[3] it will be sufficient to state here that the procedure is, briefly, as follows: (1) ions formed by collision between the bombarding electrons and the vapor or gas are observed and identified; (2) their intensities and appearance potentials, which are really upper limits, are measured; (3) these data are combined with all available thermochemical and band spectroscopic data in an effort to ascertain the processes occurring and to estimate various bond energies.

While both the H_2O and NH_3 molecules have been investigated previously,[5-7] lack of high resolving power, on the one hand, and relatively low sensitivity of the instruments on the other, have left the data on these molecules in a somewhat uncertain state. Further, more recent but incomplete data[8, 9] obtained at various times have not been in agreement with the earlier work. It appeared justifiable to presume, therefore, that a thorough investigation of the two molecules with the present apparatus might yield consistent and more complete information.

APPARATUS

The mass spectrometer which was used in the present investigation was employed previously in a study of the dissociation products of C_6H_6, C_2H_4, and HCN.[4, 10] The differential pumping system has been improved in order to make more certain that products of thermal dissociation do

---

[1] H. D. Smyth, Rev. Mod. Phys. **3**, 347 (1931). This article contains a summary of, and references to, earlier work.
[2] de Groot and Penning, *Handbuch der Physik*, Vol. 23 (Springer, Berlin, 1933). Also a resumé.
[3] L. G. Smith, Phys. Rev. **51**, 263 (1937).
[4] A. Hustrulid, P. Kusch and J. T. Tate, Phys. Rev. **54**, 1037 (1938).
[5] H. A. Barton and J. H. Bartlett, Phys. Rev. **31**, 822 (1928), on water vapor.
[6] J. H. Bartlett, Phys. Rev. **33**, 169 (1929), on ammonia.
[7] H. D. Smyth and D. W. Mueller, Phys. Rev. **43**, 116 (1933), on water vapor.
[8] P. Kusch, unpublished, on water vapor.
[9] R. F. Baker, unpublished, on ammonia.
[10] P. Kusch, A. Hustrulid and J. T. Tate, Phys. Rev. **52**, 843 (1937).

not diffuse back into the ionizing region. The filament has been moved farther from the collision chamber and the cross section of the pumping lead increased. A diagram of apparatus is given in Fig. 1. That products of thermal dissociation may readily occur is indicated by the fact that the relative abundances of the different ions are affected in a marked way by the hot filament of an ionization gauge. Since the results obtained with a tungsten filament operated very "hot" and an oxide-coated filament operated "cold" were the same, it is evident that the differential pumping was successful.

Water Vapor

Water which was distilled several times in a vacuum, and from which the middle fraction was taken, was admitted to the mass spectrometer through a long capillary. The pressure of the water vapor in the ionization chamber could be readily adjusted by varying the temperature of the bulb containing the water.

A typical mass spectrum of the heavier positive ions is given in Fig. 2. Argon was used to calibrate the mass and voltage scales. All the ions found, their relative abundance at electron energy of 100 volts, and their appearance potentials are given in Table I. On the basis of $H_2O^+=100$, the relative abundances of the impurities present were as follows: CO_2 less than 0.07; CO less than 0.05; O_2 none; and H_2 none.

Bichowsky and Rossini[11] give 2.51 volts as the energy required to dissociate one molecule of water into its components in the state in which they exist under ordinary conditions. By use of the dissociation energies of oxygen, hydrogen, and (OH), $D(O_2)=5.09$ volts, $D(H_2)=4.46$ volts, and $D(OH)=4.4$ volts given by Sponer,[12] the energies required for the dissociation of water into various components may be calculated. The values are given in Table II.

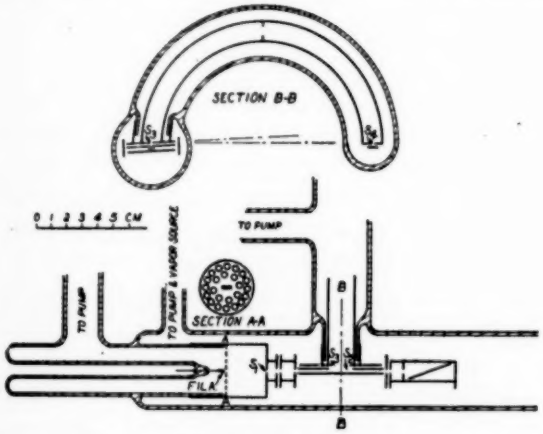


Fig. 1. Mass spectrometer with differential pumping on both analyzer and filament.

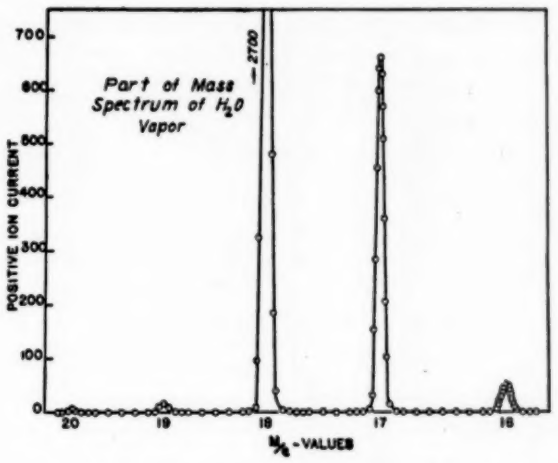


Fig. 2. Mass spectrum of heavy ion group observed in water vapor.

Positive Ions

H_2O^+

The observed ionization potential of water was 13.0 ± 0.2 volts. Barton and Bartlett[5] and Mackay[13] have observed the values 13.0 and 13.2 volts, respectively. Kusch[8] and Baker[9] obtained 12.8 and 13.0 volts, respectively. Smith and Bleakney[14] reported a value of 12.59 volts. Smyth and Mueller[7] have determined an ionization potential at 12.7 volts. Another ionization potential at 16.0 volts reported by them was not observed in the present investigation. As shown by Fig. 3, no irregularity occurred in the ionization efficiency curve from which the ionization potentials are determined. The marked change in slope observed by Smyth and Mueller certainly does not occur.

Smyth and Stueckelberg[15] have suggested that

[11] Bichowsky and Rossini, *Thermochemistry of Chemical Substances* (Reinhold, 1936).

[12] H. Sponer, *Molekülspektren* (J. Springer, 1935).

[13] C. A. Mackay, Phys. Rev. **24**, 319 (1924).

[14] L. G. Smith and W. Bleakney, Phys. Rev. **49**, 883A (1936).

[15] H. D. Smyth and E. C. G. Stueckelberg, Phys. Rev. **32**, 779 (1928).

TABLE I. *Ions observed in water vapor. Electron affinity of oxygen taken as 2.2; of hydrogen as 0.7 volt.*

Ion	Intensity at 100 Volts	Appearance Potential (Volts)	Probable Process	Calculated Minimum Energy	Excess Energy or Deduced I.P.
H_2O^+	100	13.0±0.2	$H_2O \rightarrow H_2O^+$		$I(H_2O) = 13.0$ ev
OH^+	23.2	18.7±0.2	$H_2O \rightarrow H+OH^+$	$I(OH)+$ 4.91 ev	$I(OH) \leqq 13.8$
O^+	2.0	18.8±0.5	$H_2O \rightarrow H_2+O^+$	18.6	
		28.1±1.0	$H_2O \rightarrow 2H+O^+$	23.1	$W(O^+) \leqq$ 5.0
H_2O^+		13.8±0.5			
H^+	5.0	19.5±0.2	$H_2O \rightarrow OH+H^+$	18.4	$W(H^+) \leqq$ 1.1
H_2^+	0.07	23.0±2.0	$H_2O \rightarrow O+H_2^+$	20.6	
H_1^-	0.6*	5.6±0.5	$H_2O \rightarrow OH+H^-$	4.7	
			$H_2O \rightarrow O+H+H^-$	8.8	
O^-	0.15	7.5±0.3	$H_2O \rightarrow 2H+O^-$	7.3	
		23.7±0.5	$H_2O \rightarrow H+H^++O^-$	20.8	
		36.0±3	$H_2O \rightarrow H^++H^++O^-$	34.3	

* At maximum.

there is an ionization potential of water just below that of argon. They deduce this from the apparently high probability of the reaction:

$$A^+ + H_2O \rightarrow A + H_2O^+.$$

An alternative explanation suggested by them, and more likely in view of the present results, is that ionization by a collision of the second kind is accompanied by excitation to a higher vibrational state than results from ionization by electron impact. Mulliken[16] predicts that water may have ionization potentials at 16 and 17 volts, but no evidence has been found in the present work to substantiate this prediction.

OH^+

The value $A(OH^+) = 18.7 \pm 0.2$ volts obtained here is in agreement with the values obtained by Smyth and Mueller, Kusch, and Baker. Since OH can arise only from the state *b*, Table II,

$$A(OH^+) = 5.1 + I(OH) + (\text{kinetic and excitational energy}),$$

which yields

$$I(OH) \leqq 13.6 \text{ volts}.$$

O^+

This ion appears first at 18.8±0.5 volts. The process

$$H_2O \rightarrow H_2 + O^+ \qquad (1)$$

is evidently responsible, for the minimum energy required is 18.6 volts.

A break in the efficiency curve occurs at 28.1±1.0 volts, as is shown by Fig. 4. If the process

$$H_2O \rightarrow 2H + O^+ \qquad (2)$$

is assumed to take place, the minimum energy required is 23.1 volts. This leaves 5.0 volts excess energy. While it is quite likely that the neutral hydrogen atoms may take up a considerable portion of the excess as kinetic energy, there do exist excited states in O II of energies 3.3 and 5.0 volts. These are $^2D^0$ and $^2P^0$, respectively.[17] It is to be noted that the probability of formation of O^+ by process (2) is much greater than for process (1). The failure of earlier workers[7–9] to observe the break at 28 volts may be due to the fact that a small amount of CO_2 impurity smooths out the curve to such an extent that the break is not apparent. In order to ascertain whether or not *both* processes (1) and (2) occur, it was, therefore, necessary to determine how much the CO_2 impurity contributed to the O^+ efficiency curve. A short and separate study of the O^+ from CO_2 revealed that the contribution from an amount of CO_2 comparable with that found to be present in the H_2O sample is negligible and, therefore, the above processes do occur.

TABLE II. *Energies of possible states of combination in which an oxygen and two hydrogen atoms may exist.*

State	Configuration	Energy (volts)
a	H+H+O	9.5
b	H+OH	5.1
c	H_2+O	5.0
d	H_2O	0

[16] R. S. Mulliken, Phys. Rev. **40**, 56 (1932).

[17] R. F. Bacher and S. Goudsmit, *Atomic Energy States* (McGraw-Hill, 1932).

H^+

This ion appears at 19.5±0.2 volts and evidently arises from the process

$$H_2O \rightarrow OH + H^+,$$

for which the minimum energy required is 18.4 volts. A large portion of the excess is probably taken up as kinetic energy by the products of the collision.

H_2^+

The probability of formation of this ion is small, and the appearance potential is 23.0±2.0 volts. The process

$$H_2O \rightarrow O + H_2^+$$

is the only one possible and requires 20.6 volts. If H_2^+ were the result of a secondary process or if it were present as an impurity, it would be expected to appear not far from $I(H_2)=15.3$ volts. The above process evidently does occur. Smyth and Mueller observed $A(H_2^+)$ as 33.5 volts though there was evidence of H_2^+ at lower energies. If the process

$$H_2O \rightarrow H_2^+ + O^+,$$

requiring 34.1 volts, occurs, it would be very difficult to observe the break in the curve because of the low intensity of the H_2^+ peak.

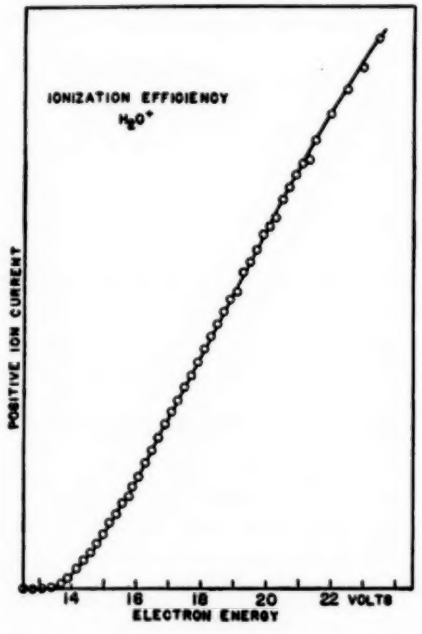


FIG. 3. Portion of the ionization efficiency curve of H_2O^+.

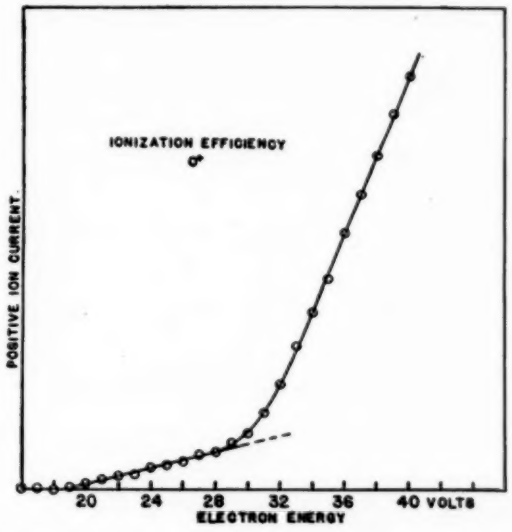


FIG. 4. Portion of the ionization efficiency curve of O^+ from H_2O.

H_3O^+

This ion appears at 13.8±0.5 volts. Its intensity relative to H_2O^+ varies as the square of the pressure. It is either the result of a secondary process or the H_3O molecule is unstable and is present in H_2O in an amount which is a function of the pressure. At about 10^{-4} mm Hg, the intensity relative to H_2O^+ is approximately 0.5.

NEGATIVE IONS

H^-

The ionization efficiency curve for H^- is shown in Fig. 5. In contrast to the curves for positive ions which rise steadily with increasing energy of the impinging electron this curve indicates that H^- is formed only when the energy is confined to a relatively narrow region. This is the well-known phenomenon of resonance capture of the electron.

H^- appears at 5.6±0.5 volts and rises to a maximum at about 7.1 volts. There is another maximum, much weaker, at about 8.9 volts which is not completely resolved from the first. Lozier,[18] and Smyth and Mueller have observed this ion and Lozier found by the method of retarding potentials that those ions forming the first maximum possess kinetic energy of 1.5 volts; those forming the second maximum possess 3.2 volts kinetic energy. The maxima which Lozier

[18] W. W. Lozier, Phys. Rev. **36**, 1417 (1930).

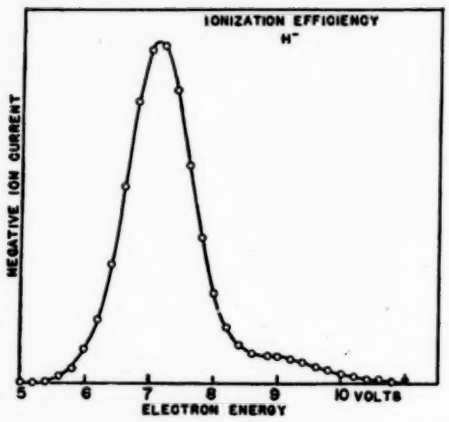


FIG. 5. Ionization efficiency curve for H^- from H_2O.

observed were centered at 6.6 and 8.8 volts, in good agreement with the present work.

The energy required to form H^- by the process

$$H_2O \rightarrow OH + H^- \qquad (1)$$

is 4.2 volts, and if we add Lozier's value of 1.5 volts kinetic energy to this, H^- should appear at 5.7 volts and this agrees well with the observed value. If it is presumed that the second group arise from process (1), but from a different excited state of the H_2O after capture of the electron, it should appear at 7.4 volts, and while this appearance is masked completely by the first group, the presence of the maximum at 8.9 volts would indicate agreement fully as good as in the preceding case.

From the present data, it would appear improbable that the process

$$H_2O \rightarrow O + H + H^-,$$

requiring 8.8 volts, occurs.

O^-

When the magnetic field of the mass spectrometer is reversed or changed for the purpose of studying ions of widely different masses, the mass scale must be recalibrated. As only one negative ion in the range M/e 18 to 16 was observed, it became necessary to introduce some gas which is known to yield a negative ion, or ions, in this range which have been identified. CO is known to yield O^-, and was therefore admitted to the apparatus, and the negative ion from H_2O, in this range, was identified as O^-.

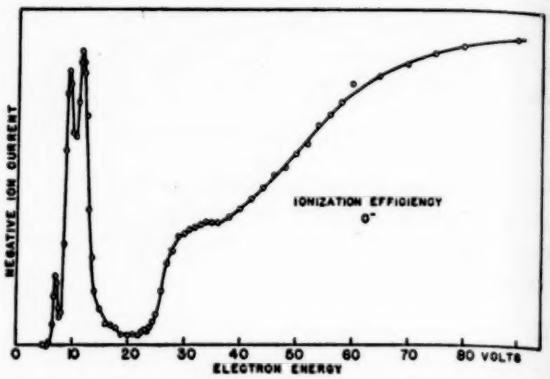


FIG. 6. Ionization efficiency curve for O^- from H_2O.

In Fig. 6 is the ionization efficiency curve of O^-. It is different from that of H^- in that three resonance peaks appear, and it is formed at all energies above the initial appearance at 7.5 ± 0.3 volts.

The process

$$H_2O \rightarrow 2H + O^- \qquad (2)$$

requires 7.3 volts, and this accounts satisfactorily for the first resonance peak. Since this peak and the second in the H^- efficiency curve occur at approximately the same energy of the bombarding electrons, it is possible that the same excited state of the H_2O^- following capture and preceding dissociation may be responsible for both. If this be so, it would appear from the relative intensities that dissociation by process (2) is more probable than by (1) from which the H^- emerges with about 3 volts kinetic energy. The second and third peaks are presumed to arise from dissociation of the H_2O from excited and unstable states following capture of the electron.

Subsequent breaks in the efficiency curve occur at 23.7 ± 0.5 volts and at 36 ± 3.0 volts. Smyth and Mueller[7] have reported these appearances as 22 ± 3 volts and 36 ± 4 volts, respectively.

They apparently result from the processes

$$H_2O \rightarrow H + H^+ + O^-, \qquad (3)$$

$$H_2O \rightarrow H^+ + H^+ + O^-. \qquad (4)$$

(3) requires 20.8 volts and (4) 34.3 volts.

A thorough search failed to reveal the presence of other negative ions. In this we have been unable to confirm the previously reported detection of OH^- in water vapor.

Ammonia

The ammonia used in this experiment was taken from a tank of liquid synthetic ammonia and fractionally distilled under vacuum. A sample of one liter at 20 cm pressure was taken from the middle portion and tests in the mass spectrometer showed that it contained less than 0.5 percent water vapor and hydrogen impurity.

Table III summarizes the results obtained. All the ions which have been observed as products of dissociation and their relative intensities at 100 volts electron energy are listed. It can be said that each ion observed is the result of single impact between an ammonia molecule and the bombarding electron for the relative intensity of each ion remained constant with variation in pressure.

In Table IV we have the energies of the several states of combination of one nitrogen and three hydrogen atoms. For the computation of these energy states, the following values of heats of dissociation and formation have been used:

$D(H_2)=4.46$ volts, $D(N_2)=7.35$ volts,

$Qf(NH_3)=0.47$ volt,[11] and $D(NH)=3.4$ volts.[19]

Positive Ions

NH_3^+

The first ionizing potential of ammonia is observed to be 10.5 ± 0.1 volts. Mackay[13] and Bartlett[6] have previously reported values of 11.1 and 11.2 volts, respectively, while Baker obtained 10.4 volts.

[19] G. Herzberg, *Molecular Spectra and Molecular Structure. I. Diatomic Molecules* (Prentice Hall, 1939).

NH_2^+

This ion can be formed by one process only, and $I(NH_2)$ is, therefore, estimated as $A(NH_2^+)-E_{\text{state } d}=11.9$ volts. The energy of the state d will be discussed under the treatment of negative ions.

NH^+

There are three processes from which this ion may originate and that at least two do occur is indicated by the shape of the ionization efficiency curve (Fig. 7). The first appearance is at 19.4 ± 0.5 volts and a sharp break in the curve is noted at 23.7 ± 0.5 volts. The difference in energy between these is 4.3 volts, a value so close to that of $D(H_2)$, that of the processes

$$NH_3\rightarrow NH^+ + H_2, \quad (1)$$

$$\rightarrow NH^+ + 2H, \quad (2)$$

$$\rightarrow NH^+ + H + H^-, \quad (3)$$

(1) and either (2) or (3) may be assigned with some confidence. $I(NH)$ is found to be $\leqq 16.4$ volts. Kusch, Hustrulid and Tate in their work on HCN find for an upper limit $I(NH)=19.2$. They took for $D(NH)$ 4.2 volts. If the value, 3.4 volts, is used, their $I(NH)$ becomes 18.4 volts,

Table IV. *Energies of the several states of combination of one nitrogen and three hydrogen atoms.*

State	Configuration	Energy (volts)
a	$N+3H$	10.9
b	$NH+2H$	7.5
c	$N+H+H_2$	6.4
d	NH_2+H	3.8
e	$NH+H_2$	3.0
f	NH_3	0

Table III. *Ions observed in ammonia.*

Ion	Intensity at 100 Volts	Appearance Potential (Volts)	Probable Process	Calculated Minimum Energy	Excess Energy or Deduced I.P.
NH_3^+	100	10.5 ± 0.1	$NH_3\rightarrow NH_3^+$		$I(NH_3)=10.5$ ev
NH_2^+	78	15.7 ± 0.1	$NH_3\rightarrow NH_2^+ + H$	$I(NH_2)+3.8$ ev(?)	$I(NH_2)=11.9$(?)
NH^+	4.5	19.4 ± 0.5	$NH_3\rightarrow NH^+ + H_2$	$I(NH)+3.0$	$I(NH)\leqq 16.4$
		23.7 ± 0.5	$NH_3\rightarrow NH^+ + 2H$	$I(NH)+7.5$	
N^+	1.7	24.9 ± 0.5	$NH_3\rightarrow N^+ + 3H$	25.4	
		28.0 ± 0.5(?)	$NH_3\rightarrow N^{+*} + 3H$	27.3	$W\leqq 0.7$
H^+	0.5	23.3 ± 0.5	$NH_3\rightarrow NH + H + H^+$	21.0	
		26.9 ± 0.5	$NH_3\rightarrow N^* + 2H + H^+$(?)	26.7	
H_2^+	0.02	15.5 ± 0.5	$H_2\rightarrow H_2^+$	15.3	
			$NH_3\rightarrow NH + H_2^+$(?)	18.3	
NH_3^{++}	0.02	42 ± 3	$NH_3\rightarrow NH_3^{++}$		$I(NH_3^+)\leqq 31.5$
H^-	0.15	5.8 ± 0.3	$NH_3\rightarrow N + H_2 + H^-$	5.7	
		23.0 ± 1.0	$NH_3\rightarrow NH^+ + H + H^-$	23.2	
NH_2^-	0.3	6.0 ± 0.5	$NH_3\rightarrow H + NH_2^-$		

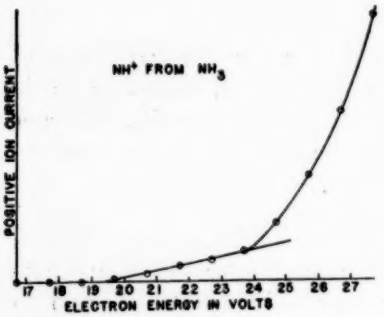


FIG. 7. Portion of the ionization efficiency curve for NH^+ from NH_3.

leaving the discrepancy at 2.0 volts. Conversely if $D(NH)=4.2$ is used in the present work $I(NH)\leqq 17.2$ volts. While it cannot be assumed flatly that in the present case the products of dissociation are all unexcited, it would appear that in the case of HCN excess energy of the order 1 volt or more does exist. The process (3) will be discussed under H^-.

N^+

This ion first appears at 24.9 ± 0.5 volts and from the table, using the energy value of state **a** and $I(N)=14.5$, we find that N^+ should occur at 25.4 volts. While these values agree within the limits of error, normally it would be expected that the experimental value lie above that calculated. That it does not is believed to be due to the fact that a small amount of N_2 impurity was found in the ammonia sample, and N^+ from N_2 is known to occur at 24.3 volts. This, then, combined with the fact that the N^+ from ammonia is of low intensity would lead to the expectation that N^+ should appear at an electron energy somewhere between the two possible values, the contribution from the impurity having the effect of masking the appearance of N^+ from ammonia very slightly.

The second appearance of N^+, at 28.0 ± 0.5 volts, while not as pronounced as the break in the NH^+ curve, does seem to exist, and it is suggested that the process

$$NH_3\rightarrow N^{+*}+3H$$

may be responsible where the excitation energy is of the order 2.5 volts. The ^1D state of N II is approximately 1.9 volts above the ground state so that this process is entirely possible.

H^+

This ion first appears at 23.3 ± 0.5 volts and from the table the energy required for the process

$$NH_3\rightarrow NH+H+H^+$$

is found to be 21 volts. This leaves 2.3 volts excess energy to be distributed among the NH and the two hydrogens and the electrons.

Energy required for the process

$$NH_3\rightarrow N+2H+H^+ \qquad (4)$$

is 24.4 volts so it is not likely that this process can account for the initial appearance of H^+. As for the break in the efficiency curve at 26.9 ± 0.5 volts, this is some 2.5 volts higher than required for the process (4), and while the nitrogen may emerge from the dissociation in an excited state, it would not be unreasonable to suppose that the hydrogens possess considerable kinetic energy. At any rate, the excess energy in both appearances of the hydrogen ion is approximately the same.

NH_3^{++}

This ion appears at 42 ± 3.0 volts, from which it is deduced that

$$I(NH^+)\leqq A(NH_3^{++})-I(NH_3)\leqq 31.5 \text{ volts.}$$

NEGATIVE IONS

NH_2^-

This ion appears at 6.0 ± 0.5 volts. The ionization efficiency curve, as shown in Fig. 8, exhibits a maximum at 7.3 ± 0.5 volts and some evidence

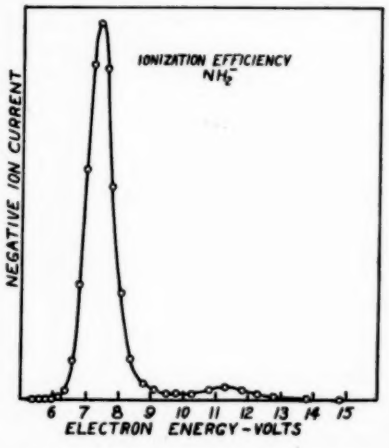


FIG. 8. Ionization efficiency curve for NH_2^- from NH_3.

of another at about 11.3 volts. At energies above 15 electron volts, this ion is not formed. The process is evidently that of resonance capture of the impinging electron.

Since this negative ion is the only one observed in the M/e range 17 to 14, it was not possible to identify it as NH_2^- until CO was admitted to the apparatus. When this was done, the peak increased in intensity, indicating the ion to be of $M/e=16$, since, as was pointed out in the discussion of water vapor, the negative ion from CO is O^-.

Of all the states in Table IV the energy of d is the most uncertain. The value, 3.8 volts, has been set more or less arbitrarily for there exists very little experimental or theoretical work from which a value of the energy necessary to remove *one* hydrogen atom from the ammonia molecule may be obtained.[20]

Since about 7.5 volts are required to remove two hydrogen atoms, it is believed that a value in the neighborhood of approximately one-half the above cannot be far wrong.

If this be so, then excess energy of the order two volts plus the electron affinity is observed in the process

$$NH_3 \rightarrow NH_2^- + H.$$

From a study of the electron configuration of the NH_2 molecule, it would be expected that it possesses characteristics somewhat similar to those of fluorine. In particular it would be expected to show a rather large electron affinity of the order 2 or 3 volts. It is, therefore, suggested that the energy of the state d is approximately correct and that NH_2^- is formed in an excited state and that the hydrogen takes up the excess energy.

H^-

This ion is also formed by resonance capture of the electron. It first appears at about 5.8 volts

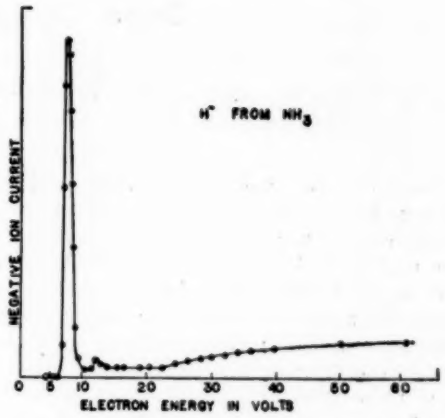


FIG. 9. Ionization efficiency curve for H^- from NH_3.

with a maximum at about 7.2 volts and another break in the current-voltage curve, indicating formation by another process, occurs at 23 volts (Fig. 9). For the initial appearance the process

$$NH_3 \rightarrow N + H_2 + H^-$$

is evidently responsible, 5.7 volts being necessary, that is, state c less 0.7 volt (electron affinity of hydrogen).

There are four possible processes which may account for the second break in the H^- curve. These are

$$NH_3 \rightarrow NH + H^+ + H^-, \quad (1)$$

$$NH_3 \rightarrow N^+ + H_2 + H^-, \quad (2)$$

$$NH_3 \rightarrow N^+ + 2H + H^-, \quad (3)$$

$$NH_3 \rightarrow NH^+ + H + H^-. \quad (4)$$

(1) requires 20.3 volts which leaves about 2.7 volts excess energy; (2) requires 20.2 volts; and (3) requires 24.7 volts. If the value of $I(NH)$ deduced above is utilized, the energy necessary for (4) is 23.2 volts. While, of course, (1) and (3) are not ruled out, (4) would appear to be the more reasonable choice even though $I(NH)$ is somewhat uncertain.

Although other negative ions such as NH^- and N^- are said to have been observed in ammonia in previous investigations, a very thorough search at electron energies from 4 volts up to 100 volts has failed to reveal any trace of these ions, and as both the resolving power and sensitivity of the present apparatus are superior with respect to the previous investigations, it can be stated with some confidence that these ions, if they do exist, are certainly of abundance <0.005.

[20] N. E. Bradbury, J. Chem. Phys. **2**, 827 (1934), in discussing data obtained by other writers on the photochemical decomposition of ammonia and his own data on the formation of negative ions in ammonia, at relatively high pressures, proposes the following processes:

$$h\nu + NH_3 \rightarrow NH_2 + H, \quad (1)$$

$$NH_3 \rightarrow NH + H_2. \quad (2)$$

The process (1) is said to occur at approximately 4.7 volts energy and (2) to require 3.0. The latter is in agreement with the energy of the state c in the table. It is not revealed how this value for (2) was arrived at.

Impact Broadening of Spectral Lines

LYMAN SPITZER, JR.
Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received May 9, 1940)

An evaluation is made of the error introduced into current impact broadening theory by the two incorrect assumptions usually made,—the adiabatic approximation, and the neglect of the rotation of the adiabatic electron states in the computation of the interaction between these states and the radiation field. A simple case of resonance broadening is examined, with the aid of wave functions in a nonrotating coordinate system, and the emitted intensity is shown to be reduced by a factor of two in the first approximation from the value given by the usual impact formula. This discrepancy is no greater than other errors in the present theory of line broadening, and will be less for dispersion, van der Waals, and other second-order forces since these forces do not vanish on the average and will show little tendency to do so in a single encounter. The Weisskopf formula, despite its theoretical imperfections, may therefore be used to give approximate results for impact broadening. A criterion is also given for the transition between impact and statistical broadening; for resonance forces under normal conditions the impact broadening formulas are shown to be irrelevant, since they apply only within a distance of 10^{-1} angstrom from the line center, and outside this range the statistical formulas must be applied. Although this is substantially the range of Doppler broadening, accurate measurements of line centers at low pressures should provide a qualitative test of the present theory in the case of resonance broadening. The influence of the previously neglected "rotation effect" on the interaction between a hydrogen atom perturbed by a passing ion and the radiation field is calculated with relative exactness for the first excited state.

THE broadening of a spectral line by perturbing atoms in a gaseous assembly has been treated in two limiting cases.[1, 2] When the atomic velocities are sufficiently small the line profile is given by the familiar statistical formulas.[3] When these velocities are large, however, the impact method of Lorentz as modified by Weisskopf[4] and others[5] is usually applied. For this analysis one computes the phase shift ϕ produced in the emitted radiation by the passing perturbing atom; if this is greater than some critical value,—unity in the Weisskopf treatment, $\pi/2$ in the Kallmann and London analysis, —the wave train before the encounter is assumed to be incoherent with that emitted after the encounter. If there are Ω such encounters per second, the line profile is then given by the usual dispersion formula $\Omega/\pi x^2$, where x equals $2\pi(\nu-\nu_0)$ and is simply the distance from the line center in angular frequency units. The term line profile is used throughout to denote $J(x)$, the profile of the line emitted by a thin film of atoms; $J(x)$ is directly proportional to the atomic absorption coefficient.

A different approach to the problem has been given by H. Kuhn,[6] who uses the first derivative of the adiabatic $\nu(t)$ to compute the time T corresponding to a phase shift π. The use of phase shifts to give line broadening directly is an approximation which may lead to serious error.[7] In any case, however, this method of calculating phase shifts is legitimate only when the total phase shift in the course of the encounter is much greater than π and the statistical theory is approximately valid. Hence this aspect of Kuhn's analysis is designed to give a minor correction to the statistical result and refers to a different region of line broadening from that of the Weisskopf impact formula in which we are primarily interested here.

Assumptions of both a mathematical and a physical nature are included in the Weisskopf analysis. From a mathematical standpoint the choice of the critical phase shift is rather arbitrary. More accurately one should compute the broadening produced by a single encounter between the radiating atom and a perturbing one; the true profile would then be given by the appropriate integration over all types of binary encounters.

While this sum could be calculated in the

[1] H. Kuhn, Phil. Mag. [7] **18**, 987 (1934).
[2] H. Margenau and W. W. Watson, Rev. Mod. Phys. **8**, 22 (1936).
[3] J. Holtsmark, Ann. d. Physik **58**, 577 (1919); Physik. Zeits. **20**, 162 (1919); **25**, 73 (1924).
[4] V. Weisskopf, Physik. Zeits. **34**, 1 (1933).
[5] W. Lenz, Zeits. f. Physik **80**, 423 (1933); H. Kallmann and F. London, Zeits. f. physik. Chemie **B2**, 207 (1929).
[6] H. Kuhn, Phil. Mag. [7] **18**, 983 (1934).
[7] L. Spitzer, Jr., Phys. Rev. **55**, 699 (1939), referred to hereafter as I.

limiting impact case, if the adiabatic approximation were assumed, such an analysis would not be very useful since the physical assumptions would not correspond to fact. In the first place the adiabatic approximation is definitely incorrect when the phase shift ϕ is as small as unity. For the adiabatic approximation breaks down when transitions between the various atomic states become appreciable. These electronic states are taken in a rotating coordinate system such that the z axis always points towards the moving perturbing atom. Transitions between atomic levels with different unperturbed energies may be neglected, except for the extremely close encounters. In the case of degenerate states of the same level, however,—states differing only in their magnetic quantum number M,—this is not true. In fact it is for the most distant encounters (the smallest values of ϕ) that such nonadiabatic transitions become important. When ϕ is large these transitions become inappreciable.

This may be shown in general by a simple dimensional argument. The change of Ψ with time will be completely determined by the time-dependent Hamiltonian. If we consider only transitions between the states of a particular atomic level, the only dimensional quantities entering the analysis are v, the velocity of the perturbing particle relative to the radiating atom; R, the distance of closest approach; q_s/r^n, the energy perturbation of state s when the perturber is stationary and at a distance r; and of course Planck's constant h. Whether or not the encounter is adiabatic must depend on the value of some dimensionless constant, and the only available one is q_s/hvR^{n-1}, which except for numerical factors of unit order of magnitude, is equal to the adiabatic phase shift ϕ_s for the state s. When v is infinite the adiabatic treatment obviously fails, and one may conclude that an encounter will be adiabatic only if ϕ_s is large compared to unity. A more specific and quantitatively more precise argument is presented in Section 1 below. Since the Weisskopf encounter radius is the value of R when ϕ_s equals unity, it is clear that the adiabatic approximation is not well suited to the determination of this radius or to the treatment of impact broadening in general.

The second assumption involved in the usual analysis of impact broadening is that the interaction between the radiation field and the adiabatically rotating electronic quantum states may be determined as though the states were not rotating. Since the radiation emitted is observed in a stationary coordinate system, one must take the matrix elements not of the electronic coordinates x, y, and z,—which are rotating about the y axis—but rather of $x \sin\theta + z\cos\theta$, y, and $z\sin\theta - x\cos\theta$, where θ is the angle which the moving z axis makes with its direction at the time t_0, when the perturbing atom is at its distance of closest approach R. It is readily shown that if the time-dependent matrix elements of these quantities are used, the perturbation produces a phase shift equal to π even when ΔE is zero or the collision is infinitely rapid. This effect is perhaps most easily seen in the light of its classical analog. A linear oscillator which is rotated through 180° will subsequently emit radiation which is out of phase by the same 180° with the previously emitted wave train, despite the absence of any perturbation to the frequency during the rotation, and despite an arbitrarily fast rotation.

This effect can obviously be taken into account only if departures from the adiabatic assumption are considered, since otherwise all encounters would be said to produce phase shifts greater than π and an enormous line broadening would be predicted. Fortunately these two deviations from the standard treatment of impact broadening tend to offset each other.

We may conclude that current pressure-broadening theory is open to serious criticism on the grounds that it assumes the adiabatic approximation, and that furthermore it does not consider the rotation effect, which should form an integral part of the adiabatic treatment. This neglect would have introduced serious error into the standard impact-broadening analyses, were it not for the fact that the adiabatic approximation was incorrect to begin with. This combination of two separate errors has happily led to results which are not far from the truth.

The simplest approach to a correct treatment of these effects is apparently to take as electronic states the eigenfunctions in a nonrotating, stationary coordinate system. Such a procedure has two advantages. In the first place it eliminates the "rotation effect"; i.e., one need consider only the usual matrix elements of the electronic co-

ordinates x, y, and z to compute the interaction with radiation. Secondly, the nonrotating unperturbed states afford a much better approximation for the case of high velocities and low phase shifts than do the rotating adiabatic states. In other words, if ϕ is as low as unity a Born-type approximation is preferable to the adiabatic one.

When the analysis is carried out for a simple case, the resultant intensity differs from the usual Weisskopf formula by a factor of two, which is scarcely more than the mathematical inaccuracy involved. But the theoretical approach required is rather different, an important fact in the theory of line broadening. While these results justify the use of the Weisskopf model for an order-of-magnitude calculation, it must be emphasized that such a model is not only mathematically approximate but is physically incorrect.

In Section 1 below the general equations for nonadiabatic transitions are examined and the region of validity for the adiabatic approximation is determined. The impact broadening formula is derived in Section 2 from the general Fourier-integral formula for $J(x)$; the conditions under which this derivation is valid lead to a simple criterion for the appearance of impact rather than statistical broadening. Section 3 presents the analysis for the broadening by resonance forces of the line from the level $L=0$ to the level $L=1$; the implications to be drawn for other cases are briefly discussed. In Section 4 a more complete treatment of the $L\alpha$ line of atomic hydrogen is given when this is perturbed by passing ions. This section provides the appropriate modifications—which qualitatively are not great—to the author's previous results,[8] in which the breakdown of the adiabatic approximation was considered, but in which the "rotation effect" was not taken into account.

1.

The general equation for the change of state of a system is

$$ih\partial\Psi(t)/\partial t=H(t)\Psi(t). \tag{1}$$

The state function $\Psi(t)$ may be expressed in terms of any complete set of eigenfunctions. Such a set is provided by the solutions of the adiabatic equation,

$$H(t)\psi_s(t)=E_s(t)\psi_s(t), \tag{2}$$

where t is simply taken as a parameter. We shall assume here and throughout that $E_s(t)$, determined by (2), is given by

$$E_s(t)=q_s/[r(t)]^n, \tag{3}$$

where $r(t)$ is the distance between the atom in question and the perturber. We shall assume for convenience that q_s is positive. Since all particles are assumed to move in straight lines, we have

$$[r(t)]^2=R^2+v^2(t-t_0)^2; \tag{4}$$

t_0 is the time of closest approach.

If we let

$$\Psi(t)=\Sigma_s a_s(t)\psi_s(t), \tag{5}$$

then for a single encounter we find from (1), (2), and (5),

$$\frac{da_s}{dt}+i\frac{E_s(t)}{\hbar}a_s+\sum_r k_{sr}a_r=0, \tag{6}$$

where

$$k_{sr}=\int\psi_s^*\frac{\partial\psi_r}{\partial t}d\tau. \tag{7}$$

Let us introduce the variable θ, defined by

$$d\theta=\frac{vdt/R}{1+v^2t^2/R^2}; \tag{8}$$

from (4) we see that θ is simply the angle which the electronic z axis makes with its position at the time of closest approach t_0.

With this change of variable (6) assumes the form

$$da_s/d\theta+i\lambda_s\delta\cos^{n-2}\theta\, a_s+\Sigma_r K_{sr}a_r=0, \tag{9}$$

where

$$K_{sr}=k_{sr}\Big/\frac{d\theta}{dt}, \tag{10}$$

$$q_s=\lambda_s q, \tag{11}$$

and

$$\delta=q/\hbar vR^{n-1}. \tag{12}$$

The quantity q is chosen so that $|\lambda_s-\lambda_r|\,(s\neq r)$ has unity as its lowest value. For most cases of importance λ_s will assume values of unit order of magnitude.

Values of K_{sr} for a rotating system have been calculated by J. Schwinger[9] for the case in which

[8] L. Spitzer, Jr., reference 7; Phys. Rev. **56**, 39 (1939), referred to here as II.

[9] J. Schwinger, Phys. Rev. **51**, 648 (1937).

the total angular momentum J is quantized. Only the rotation is relevant in the present connection, since the change in $E_s(t)$ with time gives rise only to matrix elements connecting $a_s(t)$ with states of the same M but of different unperturbed energies. As long as ΔE, the difference in unperturbed energies, is greater than hv/R, such matrix elements lead to negligible transition probabilities, and may therefore be disregarded. Since this criterion will be satisfied for most of the cases in which we are interested, we may use Schwinger's values, derived on the assumption that $E_s(t)$ is constant; from his formula (26), with ϕ set equal to zero, we find

$$K_{sr}=\tfrac{1}{2}\{(J-M)(J+M+1)\}^{\frac{1}{2}}\delta_{M,M+1} -\tfrac{1}{2}\{(J+M)(J-M+1)\}^{\frac{1}{2}}\delta_{M,M-1}, \quad (13)$$

where $\delta_{\mu\nu}$ is zero or unity as μ is different from or equal to ν. Hence $|K_{sr}|$ lies between $2^{-\frac{1}{2}}$ and $(J+1)/2$.

On the adiabatic approximation we neglect K_{sr} and solve (9) for the probability amplitude of state s; this becomes

$$a_s(t)=A_s\exp\left(-i\lambda_s\delta\int_0^{\theta(t)}\cos^{n-2}\varphi d\varphi\right), \quad (14)$$

where A_s is a constant. If we set A_s equal to unity and use (14) to calculate the first-order transition probability to another state, for which the probability amplitude was initially zero, we have from (9)

$$|a_r(\infty)|=2K_{rs} \times\left|\int_0^{\pi/2}d\theta\cos\left\{(\lambda_r-\lambda_s)\delta\int_0^{\theta}\cos^{n-2}\varphi d\varphi\right\}\right|. \quad (15)$$

It is obvious from (15) that when $|(\lambda_r-\lambda_s)\delta|$ is less than unity the argument of the cosine is always smaller than π; $|a_r(\infty)|$ may therefore be of unit order of magnitude and for such a case the adiabatic approximation is clearly invalid. When $|(\lambda_r-\lambda_s)\delta|$ is greater than K_{rs}, on the other hand, the cosine term oscillates rapidly, $|a_r(\infty)|$ is small, transitions are negligible, and the adiabatic approximation is more nearly correct. The fact that A_s, or $a_s(t_0)$, is actually less than unity does not affect these conclusions.

The adiabatic phase shift for the state s, which we shall denote by ϕ_{as}, is given by the relationship

$$\phi_{as}=\int_{-\infty}^{+\infty}\frac{2\pi E_s(t)}{h}dt=\frac{q_s}{\hbar}\int_{-\infty}^{+\infty}\frac{dt}{(R^2+v^2t^2)^{n/2}}, \quad (16)$$

provided that (3) is used for $E_s(t)$. Carrying out the integration, we find

$$\phi_{as}=c_n\lambda_s\delta, \quad (17)$$

where λ_s and δ are defined in (11) and (12), and

$$c_n=\pi^{\frac{1}{2}}\Gamma[\tfrac{1}{2}(n-1)]/\Gamma(\tfrac{1}{2}n). \quad (18)$$

Combining (17) with the results derived from (15), we see that the adiabatic approximation is valid if

$$\phi_{as}>c_nK_{sr}\left|\frac{\lambda_s}{\lambda_s-\lambda_r}\right|, \quad (19a)$$

and is wholly invalid if

$$\phi_{as}<c_n\left|\frac{\lambda_s}{\lambda_s-\lambda_r}\right|. \quad (19b)$$

Since c_n is of unit order of magnitude, we see that when ϕ_{as} is unity (19a) will not be satisfied and the adiabatic approximation definitely fails.

2.

The frequency distribution of quanta emitted by an atom in a certain excited level may be expressed as the square of a sum of the Fourier integrals of each $b_s(t)$, the probability amplitude of state s of that level;[10] one must stipulate, however, that state s is an eigenstate in a stationary coordinate system, and that the lower level of the radiative transition is undegenerate. The problem of line broadening is the evaluation of these Fourier integrals when the radiating atom is perturbed in various ways. This problem in the general case is one of great complexity and several approximations are necessary if an analysis is to be carried through.

The first simplification made in the impact case is to assume that the perturbations are very

[10] See H. Margenau and W. W. Watson, reference 2, V. Weisskopf, reference 4, and the author, I, reference 7, for proofs of this procedure under various different simplifying conditions. A general analysis for the case where both upper and lower levels are degenerate and perturbed has apparently not been given.

rapid compared to the mean life of each atomic state. Then each $b_s(t)$ may be expressed as the product of $a_s(t)$ and the radiation damping factor, where $a_s(t)$ is given by the solution of the perturbation equations, neglecting the interaction with radiation. The second approximation is to neglect all but binary encounters; this will be valid for frequencies far in the wings of the line. On this approximation the change in $a_s(t)$ at any one time is always produced by a single binary encounter; the values of a_s at the end of each encounter are taken as the initial values of a_s for the next encounter. Since the Fourier integral of $b_s(t)$ may be reduced to one of da_s/dt, this integral may be regarded as the sum of Fourier integrals, one for each separate encounter.

With these approximations, then, there are three distinct steps involved in the computation of line broadening by moving atoms. First one must set up and solve the general quantum-mechanical equation of state for the atom in the field of some passing perturber,—ion, atom, or molecule. If the state function Ψ is expanded in terms of ψ_s, a complete set of eigenfunctions, the coefficients $a_s(t)$ in the expansion will represent the solution to this first problem.

Next one must compute the Fourier integral or transform of each da_s/dt. The squared sum of these Fourier integrals over different values of s yields the broadening of the line profile for a single encounter. Finally the third step consists of integrating such profiles over encounters at all velocities and at all relevant distances of closest approach to yield the final profile $J(x)$.

If we assume, then, that the lower state is an undegenerate S state, and normalize $J(x)$ to unity for a single component of the spectral line, we have, accordingly,

$$J(x)=\frac{\Gamma}{2\pi}\sum_s\left|\int_0^\infty b_s(t)e^{ixt}dt\right|^2, \tag{20}$$

where Γ is the damping constant for the upper level (the same for all the states in a given level). The summation sign has been taken outside the brackets since the transitions from the different upper states to the lower S state will give quanta with different polarization vectors, and will therefore not interfere with each other. Following the assumption made above, we may set

$$b_s(t)=e^{-\frac{1}{2}\Gamma t}a_s(t); \tag{21}$$

$a_s(t)$ is the coefficient of ψ_s found from the perturbation equations for the encounters without regard for the interaction with the radiation field. Any perturbation to the energy of the lower state may be taken into account by a simple change in the equation for $a_s(t)$. If we substitute (21) into (20) and integrate by parts we find

$$J(x)=\frac{\Gamma}{2\pi}\frac{1}{x^2+\frac{1}{4}\Gamma^2}\sum_s\left|a_s(0)+\int_0^\infty e^{-\frac{1}{2}\Gamma t}\frac{da_s}{dt}e^{ixt}dt\right|^2. \tag{22}$$

The integral in (22) may be regarded as the sum of separate integrals, one for each encounter. If t_0 represents the midpoint in time of each encounter and if we let $\tau=t-t_0$ and neglect the damping factor inside the integral, we have

$$J(x)=\frac{\Gamma}{2\pi}\frac{1}{x^2+\frac{1}{4}\Gamma^2}\sum_s\left|a_s(0)+\sum t_0\exp\left(-\tfrac{1}{2}\Gamma t_0+ixt_0\right)\times\int_{-\infty}^{+\infty}\frac{da_s}{d\tau}(\tau)e^{ix\tau}d\tau\right|^2; \tag{23}$$

the sum over t_0 is a sum over all collisions. Since the cross-product terms vanish on the average if x is not too small, we have, finally, neglecting the radiation damping term, and integrating over t_0,

$$J(x)=\frac{1}{2\pi x^2}\sum_s\sum{}'\left|\int_{-\infty}^{+\infty}\frac{da_s}{d\tau}e^{ix\tau}d\tau\right|^2, \tag{24}$$

where Σ' denotes the sum over all encounters within an interval of one second.

One may now see the conditions under which the Weisskopf treatment is legitimate. If R represents the distance of closest approach and v the velocity of the perturbing particle relative to the atom under consideration, the encounter will last roughly $2R/v$ seconds; i.e., the value of the integral in

(24) will come largely from this range. If Rx/v is less than unity the exponential in (24) may therefore be neglected and the integral becomes simply the total change in a_s produced by the encounter.

The error introduced by this procedure may be evaluated. We introduce the dimensionless variables u and ξ, defined by

$$u=v\tau/R \tag{25}$$

$$\xi=Rx/v. \tag{26}$$

Then we find that

$$\int_{-\infty}^{+\infty}\frac{da_s}{d\tau}e^{ix\tau}dt=\int_{-\infty}^{+\infty}\frac{da_s}{du}du+\int_{-\infty}^{+\infty}(e^{i\xi u}-1)\frac{da_s}{du}du. \tag{27}$$
$$=\Delta a_s+Q.$$

From the principle of spectroscopic stability it follows that the sum over s of $|da_s/du|^2$ must equal the corresponding sum in the adiabatic case; hence we may set $|da_s/du|$ equal to its adiabatic counterpart in finding an approximate upper limit for Q.

From (9) we have in the adiabatic case

$$\frac{da_s}{du}=-\frac{i\lambda_s\delta}{(1+u^2)^{n/2}}a_s. \tag{28}$$

Taking absolute values we see from (27) that

$$|Q|\leq\int_{-\infty}^{+\infty}|e^{i\xi u}-1|\left|\frac{da_s}{du}\right|du. \tag{29}$$

If we take for $|da_s/du|$ the expression in (28) and equate $|a_s|$ to unity, (29) becomes

$$|Q|\leq 4\lambda_s\delta\int_0^\infty\frac{|\sin\tfrac{1}{2}\xi u|}{(1+u^2)^{n/2}}du. \tag{30}$$

The numerator of (30) is always less than $\frac{1}{2}\xi u$. With this substitution (30) may be integrated at once, provided that n is greater than 2. Substituting from (17) for δ and from (18) for c_n, we find

$$|Q|<K_n\phi_{as}|\xi|, \tag{31}$$

where

$$K_n=\Gamma(\tfrac{1}{2}n-1)/\pi^{\frac{1}{2}}\Gamma[\tfrac{1}{2}(n-1)]; \tag{32}$$

K_n decreases steadily from 1.00 for $n=3$ to 0.424 for $n=6$. The case $n=2$, in which we are not interested here, must be treated separately.

Formula (31) gives a useful limit on Q when $|\xi|$ is less than unity. When ϕ_{as} is sufficiently large the Lenz[11] criterion for the appearance of statistical broadening is automatically satisfied, and the usual statistical formulas may be used in place of the Fourier integral in (24). We shall neglect Q entirely in the determination of $J(x)$ and investigate later the error introduced by this approximation.

Returning, then, to (24) we see that since the maximum change in a_s is two, we may define a phase shift ϕ_s by the relationship

$$\int_{-\infty}^{+\infty}\frac{da_s}{d\tau}d\tau=\Delta a_s=2\sin\phi_s; \tag{33}$$

the ϕ_s so defined will clearly vanish for very distant encounters, and will become large for close encounters, when it will in fact equal the adiabatic phase shift ϕ_{as}. From (24) and (33) we have

$$J(x)=\frac{2}{\pi x^2}\Sigma_s\Sigma'\sin^2\phi_s. \tag{34}$$

Following the mathematics of the Weisskopf analysis, we may replace $\sin^2\phi_s$ by its average value $\frac{1}{2}$, and set

$$J(x)=(1/\pi x^2)\Sigma_s\Omega_s, \tag{35}$$

where Ω_s is the number of encounters per second for which ϕ_s is greater than unity. This approximation is possible only when n is greater than two, since otherwise the more distant encounters become more important than the closer ones.

As a rough indication of the amount by which (35) is in error we may take Q_1, the upper limit on Q given by (31) when ϕ_{as} equals unity. From (12), (17), and (26) we find that

$$Q_1=K_n\left(\frac{c_nq_s}{\hbar v^n}\right)^{1/(n-1)}|x|. \tag{36}$$

[11] W. Lenz, reference 5; also Zeits. f. Physik **83**, 139 (1933).

More accurately we should average $|\Delta a_s+Q|^2$ over all encounters to find a rigorous upper limit for the error in (35). Although the upper limit given by (31) becomes infinite as R approaches zero, it is possible to show that as R decreases, the adiabatic approximation becomes more nearly correct, and if Q_1 is less than unity the left-hand side of (27) becomes small; Q can therefore not become greater than two. Hence if we replace Q by its upper limit in (31) or by two, which ever is the smaller, the appropriate average can be roughly carried out.

The results of this calculation show that when n is equal to three, the average error in $|\Delta a_s|^2$ is not greater than five times Q_1. Since (31) gives an upper limit which is probably too great by a similar factor, we may take $2Q_1$ as the approximate error. Since also the average value of $|\Delta a_s|^2$ equals two, Q_1 becomes the relative error in $J(x)$. This procedure is approximately valid when n equals four, but will not be accurate when n is equal to or greater than five, since in such a case Q may increase very rapidly with decreasing R and the actual error of (35) may be quite large. This is particularly true if Q_1 is very small. An order of magnitude result may, however, be obtained in this manner.

Thus (36) gives an upper limit of x for which (35) is accurate to within some definite relative error. If we neglect the factor K_n, as this is of unit order of magnitude, and set Q_1 less than some preassigned relative error P, we find that this upper limit becomes

$$|\nu-\nu_0| < \frac{P}{2\pi}\left(\frac{\hbar v^n}{C_n q_s}\right)^{1/(n-1)}. \qquad (37)$$

This condition, which does not seem to have been generally realized, must be satisfied for all s if the impact broadening analysis is to have any relevance. It should be pointed out that these results depend on the assumption that ϕ_s is small when R is equal to the average interionic or interatomic distance, since otherwise ϕ_s will always be greater than unity and the limits for the sum in (35) must be changed.

Another condition which must be satisfied if (35) is to be valid is that x must be greater than Ω_s. It is evident that Ω_s plays the role of a collision damping factor, and for values of $2x$ less than the half-width $2\Omega_s$, (35) is clearly an incorrect approximation for $J(x)$. Also $1/\Omega_s$ is the difference in t_0 between encounters producing a phase shift of unity; i.e., the time elapsed between the midpoints of such encounters. If x/Ω_s is less than unity the cross products of successive integrals in (23) will include a phase factor less than unity and will not cancel out when an average is taken over neighboring values of this phase factor. This lower limit Ω_s for x will be less than the upper limit in (37) provided that ϕ_s is again less than unity when R is equal to the average distance of the nearest perturber. Hence if ϕ_s is greater than unity at this point, not only is the derivation of (35) incorrect, but there is no region of x to which it applies; in such a case $J(x)$ is in fact given by the usual statistical formula.

3.

We consider next the approximate determination of ϕ_s in the simplest case of resonance broadening, that of the $L=1$ level. Let A and B denote the two atoms in question, and 1, 2, and 3 the states with $M=1$, 0, and -1, respectively. The interaction potential V is given[12] by the formula

$$V=\frac{e^2}{r_{AB}^3}[x_Ax_B+y_Ay_B-2z_Az_B], \qquad (38)$$

provided that the coordinates are taken rotating about the y axis so that the z axis lies along the line connecting the two atoms. In stationary coordinates (38) becomes

$$V=\frac{e^2}{r_{AB}^3}[\mathbf{r}_A\cdot\mathbf{r}_B-3\{z_Az_B\sin^2\theta+x_Ax_B\cos^2\theta + (z_Ax_B+x_Az_B)\sin\theta\cos\theta\}]; \qquad (39)$$

θ, defined in (8), is identical with the quantity w used in I.

Let Λ, χ_1, χ_2, and χ_3 denote the wave functions for the ground state ($L=0$), and the three states of the upper level; each χ_s is an eigenfunction in a nonrotating, stationary coordinate system. Then for the ψ_s functions we have the usual combinations

$$\psi_s=\Lambda_A\chi_{sB}+\Lambda_B\chi_{sA}. \qquad (40)$$

The choice of sign in (40) affects this and the adiabatic analyses in the same manner, and is

[12] H. Margenau, Rev. Mod. Phys. 11, 1 (1939).

therefore immaterial in this connection. If we assume that

$$\Psi(t)=\Sigma_s a_s(t)\psi_s, \tag{41}$$

and make the substitutions

$$\begin{aligned} a_5&=(a_1-a_3)/\sqrt{2},\\ a_6&=(a_1+a_3)/\sqrt{2}, \end{aligned} \tag{42}$$

then Eq. (1) yields the formulas

$$\frac{1}{\cos\theta}\frac{da_5}{d\theta}+2i\delta a_5=0, \tag{43}$$

$$\frac{1}{\cos\theta}\frac{da_6}{d\theta}-i\delta a_6(1+3\cos 2\theta)+i\delta a_2\sin 2\theta=0, \tag{44}$$

$$\frac{1}{\cos\theta}\frac{da_2}{d\theta}-i\delta a_2(1-3\cos 2\theta)+i\delta a_6\sin 2\theta=0, \tag{45}$$

where

$$\delta=e^2\bar{r}^2/\hbar vR^2; \tag{46}$$

$\bar{r}$ is the matrix element of r between the $L=0$ and the $L=1$ quantum states.

The state represented by a_5 follows the adiabatic equation rigorously. For states a_2 and a_6 we may find a first approximation for the change in $a(t)$ as θ increases from $-\pi/2$ to $+\pi/2$. Let a_{s0} denote the initial value of a_s; remembering (33), we see that when the change in a_s is small, $\sin\phi_s$ may be replaced by ϕ_s. A simple integration yields

$$\phi_5=2i\delta a_{50};\quad \phi_6=2i\delta a_{60};\quad \phi_2=0. \tag{47}$$

Since the number of encounters per second closer than a fixed R is proportional to R^2 it is also proportional to $1/\delta$, as we see from (46). Hence if we substitute for δ from (47), setting each ϕ_s equal to unity, and equating a_{50} to a_{60}, we see that

$$\Sigma_s\Omega_s=(2+2+0)K=4K, \tag{48}$$

where K is a constant.

On the adiabatic approximation, however, one finds to the same approximation

$$\Sigma_s\Omega_s=(2+2+4)K=8K. \tag{49}$$

We see that $J(x)$ will be less than calculated on the adiabatic approximation by a factor of two. This error is not very large compared to the other uncertainties in the analysis. This decreased value of $J(x)$ arises directly from the fact that resonance forces are zero on the average and tend accordingly to cancel out even in a single encounter. A similar effect of comparable magnitude has been pointed out in II for the hydrogen lines, where again the perturbations occur in the first approximation. The neglect of the more distant collisions, which for hydrogen can become very important, is not serious here.

Higher exponents in the perturbation formula q_s/r^n correspond primarily to second approximations, and give perturbations which do not cancel out on the average. For such perturbations, in fact, the average algebraic value of the perturbation energy, taken over all states, should be approximately equal to the root mean square value. One may therefore expect the Weisskopf formula to hold more accurately for dispersion and van der Waals forces than for resonance ones.

A comparison of theory with experiment would of course be desirable. A lack of accurate knowledge of the matrix elements involved makes such a check difficult for the most part. In the case of resonance forces, however, a comparison is ideally possible. It is well known that in such a case the statistical and the adiabatic impact formulas lead to essentially the same result, independently of the velocity.[13] Hence the true impact dispersion curve should lie somewhat below the statistical one.

Unfortunately the statistical formula for this case holds even when $|\nu-\nu_0|$ is very small, corresponding to a wave-length difference of about a tenth of an angstrom. If we set q equal to $e^2hf/4\pi m\nu_0$ in (37) and assume that v does not exceed 10^5 cm/sec., we find that $|\Delta\lambda|$ must not exceed $0.30Pf^{-\frac{1}{3}}$ angstrom unit or the K resonance line, if the broadening is to follow the impact laws with a relative error not greater than P. Since the oscillator strength f is nearly unity for such lines it is clear that impact broadening is relevant only to the core of the line; i.e., roughly a tenth of an angstrom on each side of the center. Since the width produced by Doppler broadening is of the same order of magnitude, an experimental test of the present theory would be difficult but not impossible. Accurate observations with high dispersion should show significant deviations from the normal line con-

[13] H. Margenau and W. W. Watson, reference 2, pp. 41, 43.

tour given by integrating a dispersion curve over a Doppler curve. The measurements would have to be carried out at low pressures to eliminate the effect of multiple encounters which would otherwise be important in the line core.

4.

In the hydrogenic case an exact solution may be given for the profile of the line perturbed by a single encounter between the atom and an ion. The analysis has, therefore, considerable theoretical interest. In previous papers[7, 8] by the author the "rotation effect" was not considered. The results in I are correct for large and small ϕ, however, since when δ is small the equations derived for a stationary coordinate system in Section 3-I provide a check on the formulas used, while for large δ the statistical theory is valid. For intermediate values of δ, on the other hand, the results must be slightly changed.

Let us consider as before the first excited state of the hydrogen atom. Following the notation of Section 2-I we may let ψ_1, ψ_2, ψ_3, and ψ_4 represent the states in which k_1, k_2, and m are equal to 100, 001, 00–1, and 010, respectively. These functions are taken in a rotating coordinate system and are the eigenfunctions on the adiabatic approximation. The solution of Eqs. (46-I) is still valid, and shows that the a_s have the form $\exp i\sigma\theta$, $\exp -i\sigma\theta$, and a constant, where

$$\sigma^2=1+\delta^2 \tag{50}$$

and

$$\delta=-\frac{e^2z_{11}}{\hbar vR}=\frac{\phi_1}{\pi}. \tag{51}$$

More specifically the complete Ψ function may be expressed in the form

$$\begin{aligned}\Psi=&A_1\times(e^{i\sigma\theta}/2\sigma)\left\{\frac{i}{\sigma-\delta}\psi_1-\frac{i}{\sigma+\delta}\psi_4+\psi_2+\psi_3\right\}\\&+A_2\times 2^{-\frac{1}{2}}\{\psi_2-\psi_3\}\\&+A_3\times(2^{-\frac{1}{2}}/\sigma)\{i(\psi_1+\psi_4)-\delta(\psi_2+\psi_3)\}\\&+A_4\times(e^{-i\sigma\theta}/2\sigma)\left\{-\frac{i}{\sigma-\delta}\psi_1\right.\\&\qquad\left.+\frac{i}{\sigma+\delta}\psi_4+\psi_2+\psi_3,\right\},\end{aligned} \tag{52}$$

where the A_s are constants such that the sum of the squares of their absolute values is unity. One must next take the matrix components of $i(z+ix)e^{i\theta}$, $i(z-ix)e^{-i\theta}$, and y, which equal $z+ix$, $z-ix$, and y, respectively, in stationary coordinates. The line profile may be computed in a manner similar to that used in I, and in place of (20-I) and (21-I), we find

$$J(x)=\frac{\Gamma}{2\pi x^2}\left\{6+\frac{\Omega(\delta,\xi)}{\Gamma}\delta^2\{g_\sigma^2(\xi)+g_\sigma^2(-\xi)\}\right\}, \tag{53}$$

where ξ equals xR/v and

$$g_\sigma^2(\xi)=\frac{\delta^2\{\frac{1}{2}f_{\sigma+1}^2(\xi)+\frac{1}{2}f_{\sigma-1}^2(\xi)\}+f_1^2(\xi)}{\delta^2+1}; \tag{54}$$

and, as before,

$$f_\delta(\xi)=\int_{-\infty}^{+\infty}\frac{e^{i(\xi u-\delta\tan^{-1}u)}}{1+u^2}du. \tag{55}$$

Other symbols have the same meaning as in I. Formula (53) includes the effect of all four components of $L\alpha$; if (21-I), which applies to only a single component, is summed over all four, a formula identical with (53) is found, except that $g_\sigma(\xi)$ is replaced by $f_\delta(\xi)$ or, in accordance with Section 2-I, by $f_\sigma(\xi)$. The function $f_\delta(\xi)$ is shown graphically in Fig. 1 of I.

The rotation effect manifests itself here in two ways. In the first place it replaces $f_\sigma^2(\xi)$ by the average of $f_{\sigma+1}^2(\xi)$ and $f_{\sigma-1}^2(\xi)$. In the second place it introduces the term $f_1^2(\xi)/\sigma^2$, arising from the state function of which A_3 is the coefficient in (52). This state is completely unperturbed in the rotating coordinate system. Yet it is responsible for all the impact broadening when δ is small, and for $\frac{2}{3}$ of it when δ is greater than unity! This case is unusual, however, and can arise only when most of the transitions between two levels are between such "unperturbed" states,—in this case such transitions make up $\frac{2}{3}$ the total number. Formulas (53) and (54) may also be derived, of course, from a consideration of the functions in a stationary coordinate system.

The replacement of $f_\sigma(\xi)$ by $g_\sigma(\xi)$ decreases the contributions of states A_1 and A_4 to the line profile, however, almost as much as it increases that of state A_3; as a result the observable profiles

given in II are changed only in detail. This result arises from the fact that $g_\infty(\xi)$ and $g_1(\xi)$ are equal to $f_\infty(\xi)$ and $f_1(\xi)$, respectively. For intermediate values of σ the difference between $f_\sigma(\xi)$ and $g_\sigma(\xi)$ will not be very large, and we may as before take for $L(x, u)$ the two asymptotic forms; these now become

$$L(x, u)=\pi u^{\frac{1}{2}}/\gamma_1^{\frac{1}{2}} \tag{56}$$

for large u,—the same as (22a-II),—and for small u,

$$L(x, u)=4 \ln u/\gamma_1+3.77. \tag{57}$$

This substitution of 3.77 for -0.69 in the asymptotic form of $L(x, u)$ changes somewhat the final results for $H'(x)$. In particular the values of $H(x)$ when γ_1 is of unit order of magnitude or when γ_2 is only slightly less than unity may differ considerably from those given in II. The asymptotic results portrayed in Fig. 2–II, however, should still be correct within at least twenty percent.

These results are valid for the lowest Lyman line. For higher lines a separate investigation would be required for rigorous results. The main features of Fig. 2-II depend primarily, however, on the two asymptotic results,—for δ large and for δ small,—which are known to be correct for all the hydrogen lines. The results of II, summed over all components of each line, may therefore be applied to all hydrogen or hydrogenic lines.

The author is much indebted to Dr. Henry Margenau for several very valuable discussions.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

The Phenomenological Theory of Superconductors

EMORY COOK
Ryerson Physical Laboratory, The University of Chicago, Chicago, Illinois

(Received June 6, 1940)

Proceeding from the idea that the superconducting electrons are free, classical particles under the influence solely of the electric field, London found it necessary to make an integration constant vanish arbitrarily in order to obtain agreement with the Meisner effect. A variation principle constructed from the same classical ideas gives the experimental result without resorting to an additional assumption. The very precise, linear approximation results in London's equations; the rigorous theory has the same qualitative character. Unlike Mie's theory, the present theory has a satisfactory gauge invariance.

THE LONDON THEORY

THE first attempt to formulate a system of equations for a superconducting medium proceeded from the assumption that the conduction electrons are free, classical particles under the influence solely of the electric field. According to this idea, the Maxwell equations for the fields are to be supplemented by the Newtonian force equation for each electron. It is assumed that the dielectric constant and the magnetic susceptibility are unity since this is very nearly true for the metals in their normally conducting states. The equations are then:

$$\nabla\times\mathbf{H}-\mathbf{E}'=\mathbf{J}, \quad \nabla\cdot\mathbf{E}=\rho, \tag{1}$$

$$\nabla\times\mathbf{E}+\mathbf{H}'=0, \quad \nabla\cdot\mathbf{H}=0, \tag{2}$$

$$m(dVi/dt)=-e\boldsymbol{E}. \tag{3}$$

The summation of the Newtonian Eqs. (3) over the N conduction electrons per unit volume gives the macroscopic equation

$$\mathbf{E}-\lambda\mathbf{J}'=0, \quad \lambda=\frac{1}{N}\cdot\frac{m}{e^2}=\left(\frac{10^{23}}{N}\right)\cdot 10^{-11}\ \text{cm}^2. \tag{4}$$

This equation is in agreement with the experimental finding that e.m.f.'s are required not to maintain a current but only to change one. Its substitution in the second Maxwell equation results in

$$\frac{\partial}{\partial t}(\mathbf{H}+\lambda\nabla\times\mathbf{J})=0 \quad \text{or} \quad \mathbf{H}+\lambda\nabla\times\mathbf{J}=\mathbf{H}_0(x), \tag{5}$$

where $\mathbf{H}_0(x)$ is independent of the time. Now, since λ is very small, $\mathbf{H}(x, t)-\mathbf{H}_0(x)$ is practically

zero (unless the derivatives of $\mathbf{J}$ are extremely large) so that a field $\mathbf{H}_0(x)$, present in the medium when it became superconducting, is frozen in for all time. But an internal field different from zero is contrary to the Meisner effect. The normal component of the magnetic field on the surface of an ideal superconductor is always zero; no penetration of the external field is ever observed. To obtain this experimental result, London[1] arbitrarily assumed that the integration constant vanishes,

$$\mathbf{H}_0(x)\equiv 0 \quad \text{(EMPIRICAL)},$$

and based his theory on the stronger magnetic condition

$$\mathbf{H}+\lambda\nabla\times\mathbf{J}=0 \tag{6}$$

together with the acceleration condition (4) and the Maxwell Eqs. (1) and (2).[2]

A Variation Principle

Eckart[3] has proposed a variation principle as a rather general scheme for electrodynamic theories. The general theory contains the Maxwell equations, and the Lagrangian provides a convenient means for incorporating further assumptions about the particular medium. A Lagrangian constructed from the same classical ideas which London employed gives the result, which he was forced to assume, that the integration constant $\mathbf{H}_0(x)$ is identically zero.

Since the superconducting electrons are considered as classical, non interacting particles, their contribution to the Lagrangian is simply their total kinetic energy, $\frac{1}{2}NmV^2$, N being the numerical density of electrons and $\mathbf{V}$ their average velocity. The contribution of the fields is the familiar term $\frac{1}{2}(\mathbf{H}^2-\mathbf{E}^2)$. The first pair of Maxwell equations is then imposed as a side condition, giving the variation principle[4]

$$\delta\int[\tfrac{1}{2}(\mathbf{H}^2-\mathbf{E}^2)-\tfrac{1}{2}Nm\mathbf{V}^2]d\tau dt=0,$$

$$\nabla\times\mathbf{H}-\mathbf{E}-\mathbf{J}=0, \quad \nabla\cdot\mathbf{E}-\rho=0, \tag{7}$$

$$\mathbf{J}\equiv e(N_0\mathbf{V}_0-N\mathbf{V}), \quad \rho\equiv e(N_0-N).$$

Here N_0, the density of positive ions in the ionic lattice, is constant; $\mathbf{V}_0$, the velocity of the conductor itself, is zero when it is at rest. Only $\mathbf{V}_0=0$ will be considered.

1. The Eulerian equations

The electrodynamic description of the medium is to be furnished by the Eulerian equations of the above variation problem. The multiplier rule gives these equations in terms of derivatives of the function

$$\mathfrak{L}=\tfrac{1}{2}(\mathbf{H}^2-\mathbf{E}^2)+\tfrac{1}{2}Nm\mathbf{V}^2 + \varphi(\nabla\cdot\mathbf{E}-\rho)-\mathbf{A}\cdot(\nabla\times\mathbf{H}-\mathbf{E}'-\mathbf{J}),$$

where the functions φ and $\mathbf{A}$ are the Lagrangian multipliers associated with the side condition (7). It will appear that these multipliers are the electromagnetic potentials.

The conditions arising from the variation of E, H, N, V are, respectively,

$$\mathbf{E}=-\nabla\varphi-\mathbf{A}', \tag{8}$$

$$\mathbf{H}=\nabla\times\mathbf{A}, \tag{9}$$

$$\tfrac{1}{2}m\mathbf{V}^2+e\varphi-e\mathbf{A}\cdot\mathbf{V}=0, \tag{10'}$$

$$m\mathbf{V}-e\mathbf{A}=0. \tag{11}$$

The combination of Eqs. (10') and (11) gives the more compact relation:

$$\tfrac{1}{2}m\mathbf{V}^2-e\varphi=0. \tag{10}$$

Equations (7–11) may be combined to give the energy and momentum equations:

$$\frac{\partial}{\partial t}[\mathbf{H}^2+\mathbf{E}^2+\tfrac{1}{2}Nm\mathbf{V}^2] +\nabla\cdot[\mathbf{E}\times\mathbf{H}+\tfrac{1}{2}Nm\mathbf{V}^2\mathbf{V}]=0, \tag{12}$$

$$\frac{\partial}{\partial t}[\mathbf{E}\times\mathbf{H}+\mathbf{N}m\mathbf{V}]+\tfrac{1}{2}\nabla[\mathbf{H}^2+\mathbf{E}^2] -\nabla\cdot[\mathbf{EE}+\mathbf{HH}-Nm\mathbf{VV}]=-eN_0\mathbf{E}, \tag{13}$$

which have precisely the terms to be expected for a classical electronic gas that interacts with the positive ionic lattice only through the electric field.

[1] F. and H. London, Proc. Roy. Soc. **A149**, 71 (1935).

[2] In the previous reference, the Londons considered various functions for the right-hand side of Eq. (4); but in later papers [F. London, Nature **140**, 706 (1937)] they returned to Eq. (4). This simple form has a direct confirmation in that a condenser exhibits no change in capacity when its plates become superconducting [F. London, Proc. Roy. Soc. **A155**, 102 (1937)].

[3] Carl Eckart, Phys. Rev. **54**, 920 (1938).

[4] The term $\Sigma_n\theta_n[\dot{N}_n+\nabla\cdot(N_n\mathbf{V}_n)]$ in Eckart's Lagrangian is an expression of the rate of production of the various types of particles. It does not occur when only one type is considered, for the first pair of Maxwell equations insures the conservation of charge which is then equivalent to the conservation of particles.

Equations (8), (9) give the usual expressions for the fields in terms of the potentials φ and $\mathbf{A}$, thus insuring the second pair of Maxwell equations. Equations (10), (11) constitute our extension of the Maxwell system and are to be compared with the Eqs. (4), (6) of the London theory. It will be shown that they are very similar.

2. The approximate, linear theory

Equations (7–11) are difficult to handle. A linear system will now be developed which approximates them very closely. From Eqs. (8–11) there results:

$$\mathbf{E}=(m/2e)\nabla\mathbf{V}^2-(m/e)\mathbf{V}', \tag{14}$$

$$\mathbf{H}=(m/e)\nabla\times\mathbf{V}, \tag{15}$$

which may be used to eliminate $\mathbf{E}$, $\mathbf{H}$ from the set (7),

$$\begin{gathered}\frac{m}{e}\nabla\times(\nabla\times V)+\frac{m}{2e}\nabla(V^2)'+\frac{m}{e}V''+eNV=0,\\ -\frac{m}{2e}\nabla^2V^2-\frac{m}{e}\nabla\cdot V'=-e(N_0-N).\end{gathered} \tag{16}$$

The notation

$$N=N_0(1+n),\quad \beta^2=(e^2/m)N_0=1/\lambda,$$

permits the simplification:

$$\begin{gathered}\nabla\times(\nabla\times\mathbf{V})+\beta^2\mathbf{V}+\mathbf{V}''=-\tfrac{1}{2}\nabla(\mathbf{V}^2)'-n\beta^2\mathbf{V},\\ n\beta^2=\nabla\cdot\mathbf{V}'+\tfrac{1}{2}\nabla^2\mathbf{V}^2.\end{gathered} \tag{17}$$

So to terms linear in n, $\mathbf{V}$ Eqs. (7–11) reduce to

$$\nabla\times(\nabla\times\mathbf{V})+\beta^2\mathbf{V}+\mathbf{V}''=0$$

or

$$\nabla\times(\nabla\times\mathbf{J})+\beta^2\mathbf{J}+\mathbf{J}''=0, \tag{18}$$

$$\mathbf{E}=-(m/e)\mathbf{V}' \quad \text{or} \quad \mathbf{E}-\lambda\mathbf{J}'=0, \tag{19}$$

$$\mathbf{H}=(m/e)\nabla\times\mathbf{V} \quad \text{or} \quad \mathbf{H}+\lambda\nabla\times\mathbf{J}=0. \tag{20}$$

This linear approximation is very precise in any physical problem because n, and $\mathbf{V}$ are very small. The charge density is enormous ($\rho=-eN_0$) when $n=1$, and each electron has the velocity of light when $\mathbf{V}=1$. Equation (19) is the acceleration condition (4) which gives the zero resistance found experimentally. Equation (20) is the magnetic relation (6) with the integration constant, $\mathbf{H}_0(x)$, identically zero as required by the Meisner effect. The linear approximation of our theory is thus equivalent to London's theory.

The economy of the variation principle is striking. London could not obtain the magnetic relation (20) from the acceleration argument alone. He was forced to make the assumption that the integration constant $\mathbf{H}_0(x)$, representing the field in the medium when it became superconducting, is always zero. The same ideas in a Lagrangian form give the experimental result without the need of a further assumption.

It is highly unlikely that there is a mistake in the acceleration argument whose correction would give $\mathbf{H}_0=0$, as in the present theory. Eckart's general variation principle must contain an assumption in addition to those of Maxwellian electrodynamics. The physical assumptions involved in the choice of the particular Lagrangian are quite clear, at least in the present simple case; they are almost identical with those involved in the choice of a Lagrangian in the mechanics of a single particle. Thus it seems that the clarification of the new assumption can come only from a physical interpretation of the theory of Lagrangian multipliers. But such an interpretation is not easily given, particularly since the mathematicians have not yet developed a satisfactory multiplier theory for multiple integrals.

3. The potentials and gauge invariance

The formulae (8), (9) justify our speaking of the Lagrangian multipliers φ and $\mathbf{A}$ as potentials; but it should be emphasized that they are a very special choice from the class of functions acceptable as potentials. They do not, in general, satisfy the Maxwell auxiliary equation; but this is not disappointing. The equation was imposed by Maxwell as a voluntary restriction on the generality of the potentials which he was to consider, his one reason for staying within this restricted domain being the simplification of the mathematical theory resulting from the separability of the equations. The present choice is more convenient for our purposes.

Our restrictions, however, are more stringent than Maxwell's. Indeed our potentials are uniquely determined by Eqs. (10), (11) for each physical situation in which $\mathbf{E}$, $\mathbf{H}$, $\mathbf{J}$ and ρ are known; whereas there is a wide variety in the Maxwellian potentials for a given problem. Their

character is apparent from the energy Eq. (10)[5] which has the right-hand side equal to zero rather than to an arbitrary constant and from Eq. (11) which gives a meaning to the absolute value of **A**.

Weyl's objection to Mie's theory was precisely the uniqueness of the potentials. He found it inconceivable that the equations of motion should depend upon their absolute value. But the objection loses its validity in the present theory because Eqs. (10), (11) may be used to eliminate the potentials from the expressions not only for the fields but for the charge and current as well. Our restrictions are purely for mathematical convenience as was the Maxwell auxiliary equation; they do not involve a physical assumption. Indeed it is possible to branch out into the larger domain of the Maxwellian potentials φ and **A** through a gauge transformation

$$\varphi_1=\varphi+\chi; \quad \mathbf{A}_1=\mathbf{A}-\nabla\chi,$$

under which Eqs. (8), (9) are invariant, i.e.,

$$\mathbf{E}=-\nabla\varphi_1-\mathbf{A}_1', \tag{21}$$

$$\mathbf{H}=\nabla\times\mathbf{A}_1. \tag{22}$$

To show that φ_1 and A_1 have the generality of the Maxwellian potentials, it suffices to show that the function χ has the generality of the similar function in the Maxwell theory. The sole restriction on φ_1 and $\mathbf{A}_1$ is the Maxwell auxiliary equation

$$\nabla\cdot\mathbf{A}_1+\varphi_1'=0, \tag{23}$$

which becomes

$$\chi''-\nabla^2\chi=-\varphi-\nabla\cdot\mathbf{A} \tag{24}$$

with the right-hand side known uniquely. In Maxwell's theory Eq. (24) is homogeneous, the right-hand side being zero. The most general χ is $\chi=\chi_p+\chi_0$ where χ_0 is the general solution of the homogeneous equation and χ_p is a particular solution of the complete equation. Thus χ has the generality of χ_0, proving that φ_1 and $\mathbf{A}_1$ actually are the Maxwellian potentials.

4. The rigorous theory

The higher order terms which were neglected in obtaining London's equations do not introduce even a small resistance. The rigorous equations have static solutions (with $\mathbf{J}\neq 0$), so the direct-current resistance must be zero.

[5] An expression for the energy density which is independent of the potentials was given in Section 3.

The rigorous theory is contained in the following for the parameter value $\epsilon=1$: the equation

$$\nabla\times(\nabla\times\mathbf{V})+\beta^2\mathbf{V}+\mathbf{V}'' = -\mathbf{V}[\tfrac{1}{2}\epsilon^2\nabla^2\mathbf{V}^2+\epsilon\nabla\cdot\mathbf{V}']-\epsilon\nabla(\mathbf{V}\cdot\mathbf{V}') \tag{25}$$

in the superconducting region τ, and the equations

$$\nabla\times\mathbf{H}-\mathbf{E}'=0, \quad \nabla\cdot\mathbf{E}=0, \qquad \nabla\times\mathbf{E}+\mathbf{H}'=0, \quad \nabla\cdot\mathbf{H}=0, \tag{26}$$

in free space τ_0, connected by the boundary conditions[6]

$$\mathbf{H}=(m/e)\nabla\times\mathbf{V}, \tag{27}$$

$$\nu\times\mathbf{E}=-(m/e)\nu\times(\mathbf{V}'-\tfrac{1}{2}\epsilon\nabla\mathbf{V}^2), \tag{28}$$

across the interface, S (ν is the unit normal). If the solutions are expansible in sufficiently convergent power series

$$\mathbf{V}=\mathbf{V}_0+\mathbf{V}_1\epsilon+\mathbf{V}_2\epsilon^2+\cdots,$$

$$\mathbf{E}=\mathbf{E}_0+\mathbf{E}_1\epsilon+\mathbf{E}_2\epsilon^2+\cdots,$$

$$\mathbf{H}=\mathbf{H}_0+\mathbf{H}_1\epsilon+\mathbf{H}_2\epsilon^2+\cdots,$$

then Eqs. (25–28) divide into a number of equations through the requirement that the coefficient of each power of ϵ must vanish.

$$L(\mathbf{V}_0)+\mathbf{V}_0''=0, \quad L(V)\equiv\nabla\times(\nabla\times\mathbf{V})+\beta^2\mathbf{V}, \tag{29}$$

$$\nabla\times\mathbf{H}_0-\mathbf{E}_0'=0, \quad \nabla\cdot\mathbf{E}_0=0, \tag{30}$$

$$\mathbf{H}_0-(m/e)\nabla\times\mathbf{V}_0=0, \tag{31}$$

$$\nu\times\mathbf{E}_0+(m/e)\nu\times\mathbf{V}_0'=0, \tag{32}$$

$$L(\mathbf{V}_1)+\mathbf{V}_1''=-\mathbf{V}_0\nabla\cdot\mathbf{V}_0'-\nabla(\mathbf{V}_0\cdot\mathbf{V}_0'), \tag{33}$$

$$\nabla\times\mathbf{H}_1-\mathbf{E}_1=0, \quad \nabla\cdot\mathbf{E}_1=0, \tag{34}$$

$$\mathbf{H}_1-(m/e)\nabla\times\mathbf{V}_1=0, \tag{35}$$

$$\nu\times\mathbf{E}_1+(m/e)\nu\times\mathbf{V}_1'=-(m/2e)\nu\times\nabla\mathbf{V}_0^2. \tag{36}$$

. . . . . .

Equations (29–32), being the London conditions, have static solutions (with $\mathbf{V}\neq 0$). For such a zeroth-order solution, the inhomogeneous parts of Eqs. (33–36) are independent of the time so that static first-order corrections, $\mathbf{V}_1$, $\mathbf{H}_1$, $\mathbf{E}_1$, also

[6] This condition is the one usually assumed, the continuity of the magnetic field and of the tangential components of the electric field. The question of the boundary conditions for the London theory, i.e. $\epsilon=0$, is discussed at some length in a following paper.

exist. The application of the argument to higher and higher orders shows that solutions **V, H, E** exist for which all terms are static. Thus steady currents do not encounter even the smallest resistance.

A slight modification of this argument shows that the resistance is rigorously zero for non-steady currents as well. Equations (29–32) have simple harmonic solutions,

$$\mathbf{V}_0(x, t) = \mathbf{V}_0(x)e^{i\omega t};$$

for such solutions the inhomogeneous parts of (33–36) have factors $e^{2i\omega t}$, so that they have solutions

$$\mathbf{V}_1(x, t) = \mathbf{V}_1(x)e^{2i\omega t}$$

and in general

$$\mathbf{V}_n(x, t) = \mathbf{V}_n(x)e^{(n+1)i\omega t}.$$

Thus an alternating current or field excites only harmonics of the fundamental frequency and there are periodic solutions of the equations, which would not be the case under the working of an irreversible resistance.

The conditions upon the nth correction, $\mathbf{V}_n$, $\mathbf{H}_n$, $\mathbf{E}_n$, differ from the conditions (29–32) upon the zeroth approximation only in the inhomogeneous terms. Consequently, a uniqueness theorem, once established for the solutions of the London theory, is equally valid for the rigorous solutions of the present theory. This becomes important in view of the results of the following paper (cf. reference 8).

It is a pleasure to acknowledge the assistance of Professor Eckart, who also suggested the application of the variation principle to the problem of the superconductor.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

Complete Data and Boundary Conditions for a Superconductor

EMORY COOK
Ryerson Physical Laboratory, The University of Chicago, Chicago, Illinois
(Received June 6, 1940)

When the applied fields and the total charge on each conductor are known, the solution of the London equations for a superconductor in the steady state is completely determined by requiring the continuity of the magnetic field and of the tangential components of the electric field. This result is incompatible with the somewhat prevalent notion that the lines of current flow rotate with the conductor. The solution for a nonsteady state is uniquely determined by the additional specification of **E**, **H**, **J** throughout the system at some particular time.

THE STEADY STATE

THE purpose of the present paper is to arrive at a set of data and boundary conditions for the London theory which is at the same time mathematically complete and physically meaningful. The existence of such a set is of more than mere mathematical interest; it has the following important physical implication.

An experiment has been described[1] in which a superconducting, hollow sphere in an applied magnetic field experienced a torque when turned out of its equilibrium position. This is interpreted to mean that the currents turn with the conductor. More precisely, it means that the fields and currents in the initial and final positions are different. But if this experiment is taken at its face value, it is contrary to the present theory, for it was performed in such a manner that the fields at infinity (i.e. the applied fields) are the same in the two positions. It follows from the uniqueness theorem that the fields should therefore have been the same everywhere. The experiment thus rules out, not only this theory, but every theory which has a similar uniqueness theorem.[2] In the absence of confirmation for the experiment, it appears reasonable to assume that the sphere was not homogeneously superconducting, but had non-

[1] K. H. Onnes, Comm. Leiden Suppl., No. 50a, 8–10 (1924).

[2] In particular, it is contrary to even the rigorous form of the theory resulting from the variation principle (cf. the previous paper, Section 4).

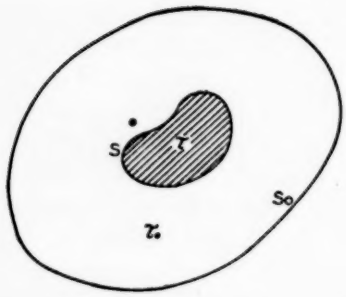


FIG. 1. A superconducting region τ, surrounded by free space τ_0, the entire system enclosed by a surface S_0.

spherical regions of ordinary conductivity in its interior.

Experience with electrodynamic theories of other types suggests that the boundary conditions across the air-metal interface are the continuity of $\mathbf{H}$ and $\nu\times\mathbf{E}$ (ν is the unit normal). The continuity of $\mathbf{H}$ is the physical assumption of unit permeability and zero surface density both of electric current and magnetic charge. The continuity of $\nu\times E$ assumes unit dielectric constant and zero surface magnetic current but admits a surface electrical charge.

The imposition of these boundary conditions still leaves many solutions corresponding to the diversity of the possible physical situations. The physical situation is completely determined by the applied fields and the total charge on each conductor.[3] It is very satisfactory that the above boundary conditions, together with the values of the charge parameters (the data) and the behavior of the solutions at infinity (the boundary data), determine the mathematical solution uniquely. In all probability these various physical conditions are mathematically independent; their independence has been established in detailed calculations for an infinite cylinder.[4]

The procedure is to express the volume integral of $E^2+H^2+\lambda J^2$ in terms[5] of integrals over a surface S_0 surrounding the system and over the air-metal interface S across which discontinuities may occur (Fig. 1). These surface integrals vanish for $\mathbf{E}$, $\mathbf{H}$, $\mathbf{J}$ which satisfy the boundary conditions and have zero data and boundary, so such $\mathbf{E}$, $\mathbf{H}$, $\mathbf{J}$ are everywhere zero. The uniqueness theorem is then immediate, for the linearity of the equations enables us to consider the identically zero solution as the difference of two solutions which satisfy the boundary conditions and have the same data and boundary data.

In discussions of the steady state, the electric and magnetic fields may be treated independently. Equation (11) of the previous paper requires that $\mathbf{E}=0$ in a superconductor just as in a normal conductor; in free space the field $\mathbf{E}$ satisfies the usual Maxwell equations. The analysis for the interaction of an electrostatic field and a system of conductors is well known[6] and need not be discussed here. The result is that the boundary data,

$$\nu\cdot\mathbf{E}, \tag{1}$$

specified on S_0, and the total charge Q,

$$Q=\int_S \nu\cdot\mathbf{E}\,dS \tag{2}$$

specified for each conductor, determine the field $\mathbf{E}$ uniquely.

The usual proof of this uniqueness of the electrostatic field depends upon the existence of a scalar potential; but a generalization of the method is applicable to the magnetic field which has a vector potential. The equations governing the magnetic field are:[7]

$$\nabla\times\mathbf{H}=0, \quad \nabla\cdot\mathbf{H}=0 \quad \text{in } \tau_0, \tag{3}$$

$$\nabla\times\mathbf{H}=\mathbf{J}, \quad \mathbf{H}=\nabla\times\mathbf{A}, \quad \mathbf{A}+\lambda\mathbf{J}=0 \quad \text{in } \tau. \tag{4}$$

Let the indices 1,2 designate different solutions of these equations, and

$$H=\mathbf{H}_1-\mathbf{H}_2, \quad J=\mathbf{J}_1-\mathbf{J}_2, \quad A=\mathbf{A}_1-\mathbf{A}_2.$$

Putting $\mathbf{P}=\mathbf{Q}=A$ in the identity[8]

$$\nabla\cdot[\mathbf{P}\times(\nabla\times\mathbf{Q})]=(\nabla\times\mathbf{P})\cdot(\nabla\times\mathbf{Q})-\mathbf{P}\cdot\nabla\times(\nabla\times\mathbf{Q})$$

gives

$$\nabla\cdot[A\times H]=H^2$$

in free space τ_0, and

$$\nabla\cdot[A\times H]=H^2+\lambda J^2$$

in the superconducting region τ. Now, since $\mathbf{J}=0$ in τ_0,

[3] Only simply connected conductors are considered here. The permanent magnetic moment of a ring-like conductor will be discussed in a future paper.

[4] Unpublished.

[5] λ is the positive constant introduced by London.

[6] Sir James Jeans, *The Mathematical Theory of Electricity and Magnetism*, fifth edition (Macmillan Company, New York), p. 163.

[7] A differs from the usual Maxwellian potential which London employs (cf. the previous paper, Section 3).

[8] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 163 (1939), have emphasized the importance of this identity.

$$\int_{\tau+\tau_0}[H^2+\lambda J^2]d\tau=\int_{S_0}\nu\cdot A\times H dS+\int_S \nu\cdot A\times H\Big]_+^- dS, \quad (5)$$

where $]_+^-$ is used to designate the difference of the values of the preceding function on the two sides of the surface. If the surface integrals in Eq. (5) vanish, then $(\mathbf{E}_1, \mathbf{H}_1)=(\mathbf{E}_2, \mathbf{H}_2)$ throughout $\tau+\tau_0$.

The integral over S_0 vanishes if either $\nu\times A=0$ or $\nu\times H=0$. The latter is the more important physical case since the absolute value of $\mathbf{A}$ has no significance in free space. Thus the specification of the boundary data, $\nu\times\mathbf{H}$, corresponding to the applied field, i.e.,

$$\nu\times\mathbf{H}_1=\nu\times\mathbf{H}_2, \quad (6)$$

causes the first of the surface integrals to vanish.

The vanishing of the integral over S is not obtained quite so simply. The following identity, in which subscripts are used to denote the evaluation of the function on the positive or negative side of S,

$$\nu\cdot A\times H\Big]_+^- = H_-\cdot\nu\times(A_- - A_+)+A_+\cdot\nu\times(H_- - H_+), \quad (7)$$

insures the vanishing of this integral, provided

$\nu\times\mathbf{A}$, and $\nu\times\mathbf{H}$ are continuous across S. (8)

But the continuity of $\nu\times\mathbf{A}$ has no immediate physical interpretation; it is necessary to investigate the extent to which requirement (8) is equivalent to the expected boundary condition:

$\nu\cdot\mathbf{H}$ and $\nu\times\mathbf{H}$ are continuous across S. (9)

The continuity of $\nu\times\mathbf{A}$ and $\nu\times\mathbf{H}$ insures the result

$$\int_S \nu\cdot A\times H\Big]_+^- dS=0 \quad (10)$$

through the vanishing of the integrand. The question now is whether or not the continuity of $\nu\cdot\mathbf{H}$ and $\nu\times\mathbf{H}$ also suffices. The answer is in the affirmative provided there exists in τ_0 a single valued and sufficiently continuous function χ such that

$$\nu\times\nabla\chi=\nu\times A\Big]_+^-. \quad (11)$$

For the two relations, $\nabla\times\mathbf{H}=0$ in τ_0, and $\mathbf{H}_+=\mathbf{H}_-$ on S, together with the previously assumed boundary data, $\nu\times H=0$ on S_0, then permit the evaluation of the integral:

$$\int_S \nu\cdot A\times H\Big]_+^- dS=\int_S \nu\cdot[(\nabla\chi)\times H]dS$$
$$=-\int_{\tau_0}\nabla\cdot[(\nabla\chi)\times H]d\tau$$
$$+\int_{S_0}\nu\cdot[(\nabla\chi)\times H]dS=0. \quad (12)$$

The question thus reduces to that of the existence of the function χ.

The function χ will now be constructed for a hollow sphere. The topological character of this construction is such that it is possible for any superconductor which can be obtained from a hollow sphere by continuous deformation, i.e. for one which is simply connected.

The notation is defined in Fig. 2. S_0 is the surface of the bounding sphere of radius r_0. S is the complete surface, inner and outer, of the hollow sphere; r_1, r_2 are its inner and outer radii. τ is the superconducting region, τ_0 is free space.

Now if S' is any part of S, and C is its contour, then the continuity of H gives

$$0=\int_{S'}\nu\cdot H\Big]_+^- dS=\int_{S'}\nu\cdot\nabla\times A\Big]_+^-=\int_C A\Big]_+^-\cdot d\mathbf{R}.$$

This enables us to define functions of the spherical coordinates (r, ϑ, φ):

$$\psi_1(\vartheta,\varphi)=\chi(r_1,\vartheta,\varphi)=\int_{0,0}^{\vartheta,\varphi}(A_- - A_+)\cdot d\mathbf{R}\Big]_{r=r_1},$$

$$\psi_2(\vartheta,\varphi)=\chi(r_2,\vartheta,\varphi)=\int_{0,0}^{\vartheta,\varphi}(A_- - A_+)\cdot d\mathbf{R}\Big]_{r=r_2},$$

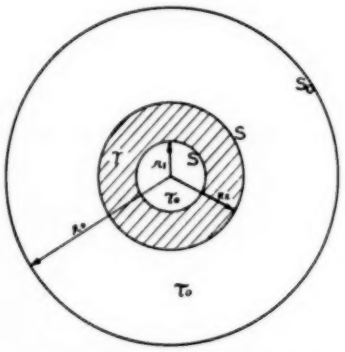


FIG. 2. A superconducting, hollow sphere τ, in free space τ_0, the entire system within a sphere of surface S_0.

which are single valued, continuous, and have the property (11).[9] There remains but to continue χ throughout the intervals $(0, r_1)$ and (r_2, r_0) so that it is defined at each point of τ_0. To this end a function $f(r)$ is introduced which is required merely to have continuous first and second derivatives and to take on the prescribed values

$$f(0)=0, \quad f(r_1)=1, \quad f(r_2)=1, \quad f(r_0) \text{ finite.}$$

The continuation

$$\chi(r, \vartheta, \varphi)=f(r)\vartheta_1(\vartheta, \varphi), \quad r \leqslant r_1,$$
$$\chi(r, \vartheta, \varphi)=f(r)\vartheta_2(\vartheta, \varphi), \quad r \geqslant r_2,$$

gives χ all the desired properties in τ_0 and on S and thus completes the proof that the conditions (6) and (9) determine the field uniquely, at least for a hollow sphere. But the spherical form of the superconductor entered only in providing a simple coordinate system for the definition of the function $\chi(r, \vartheta, \varphi)$. If the sphere is continuously deformed into an arbitrary shape, χ retains its continuity properties as well as the property (11), the only difference being that the coordinates (r, ϑ, φ) are no longer the simple spherical coordinates but are the parameters along the curves into which the original coordinate curves have been deformed. Thus conditions (6), (9) form a complete set for all simply connected superconductors.

The Nonsteady State

The general equations are

$$\left.\begin{aligned} &\nabla\times\mathbf{H}-\mathbf{E}'=0, \quad \nabla\cdot\mathbf{E}=0, \\ &\nabla\times\mathbf{E}+\mathbf{H}'=0, \quad \nabla\cdot\mathbf{H}=0, \end{aligned}\right\} \text{ in } \tau_0$$
$$\left.\begin{aligned} &\nabla\times(\nabla\times\mathbf{J})+\beta^2\mathbf{J}+\mathbf{J}''=0, \\ &\mathbf{H}+\lambda\nabla\times\mathbf{J}=0, \quad \text{in } \tau. \\ &\mathbf{E}-\lambda\mathbf{J}'=0, \end{aligned}\right\} \tag{13}$$

[9] Up to this point no use had been made of the simple connectivity of the conductor. But here it is needed to obtain the single valuedness of the ψ's. In the case of a ring-like conductor, a cut must be made to render its surface simply connected. In the subsequent integral transformations, this cut contributes an additional surface integral which vanishes only when the two solutions have the same total current around the ring. The mathematics for this case has been worked out in detail and will be presented in a future paper.

The proof of the uniqueness theorem will be based upon the energy Eq. (12) of the previous paper. Terms of order higher than the second must be discarded in the linear approximation represented by Eqs. (13). There results

$$(\partial/\partial t)[\tfrac{1}{2}(\mathbf{H}^2+\mathbf{E}^2)+\tfrac{1}{2}\lambda\mathbf{J}^2]+\nabla\cdot[\mathbf{E}\times\mathbf{H}]=0,$$

which is valid in the composite region $\tau+\tau_0$ with the understanding that $\mathbf{J}=0$ in τ_0. If script letters are again used to denote the difference of two solutions, then

$$\int_{\tau+\tau_0}\int_0^t [\tfrac{1}{2}(\mathcal{H}^2+\mathcal{E}^2)+\tfrac{1}{2}\lambda\mathcal{J}^2]dtd\tau = -\int_{S_0}\nu\cdot\mathcal{E}\times\mathcal{H}dS+\int_S \nu\cdot\mathcal{E}\times\mathcal{H}\Big]_-^+ dS,$$

which gives immediately the result:

If two solutions, together with their first derivatives, are continuous except possibly across the air-metal interface S, satisfy the boundary conditions:

$$\left.\begin{aligned} &\nu\times\mathbf{E} \text{ and } \nu\times\mathbf{H} \text{ continuous for } t\geqslant 0, \\ &\nu\cdot\mathbf{H} \text{ continuous at } t=0, \end{aligned}\right. \tag{14}$$

and have the same values for the boundary data,

$$\nu\times\mathbf{E} \text{ or } \nu\times\mathbf{H} \text{ for } t\geqslant 0, \tag{15}$$

on the bounding surface S_0, as well as for the initial data

$$\mathbf{E}, \mathbf{H}, \mathbf{J}, \text{ in } \tau+\tau_0 \text{ at } t=0, \tag{16}$$

then the solutions are identical in $\tau+\tau_0$ for $t\geqslant 0$ and $\nu\cdot\mathbf{H}$ is continuous for each.

The continuity of $\nu\cdot\mathbf{H}$ is not read out of the above integral transformation as are the other conclusions of the theorem; but is a consequence of the continuity of $\nu\times\mathbf{E}$ and the initial continuity of $\nu\cdot\mathbf{H}$. For any part S' of S,

$$\frac{\partial}{\partial t}\int_{S'}\nu\cdot\mathbf{H}\Big]_-^+ dS=-\int_{S'}\nu\cdot\nabla\times\mathbf{E}\Big]_-^+ dS = -\int_C \mathbf{E}\Big]_-^+\cdot dR=0,$$

and so $\nu\cdot\mathbf{H}]_-^+=0$ on S for $t\geqslant 0$.

Temperature Dependence of Young's Modulus of Beta-Brass Single Crystals

JOHN S. RINEHART
Department of Physics, State University of Iowa, Iowa City, Iowa
(Received June 24, 1940)

Young's modulus of seven single crystals of beta-brass has been measured as a function of temperature from 25°C to 505°C for the purpose of studying the order-disorder phenomenon. The method used is that of the composite piezoelectric oscillator. The reciprocal of Young's modulus is linearly related to the orientation function at all temperatures. The maximum and minimum values of the reciprocal of Young's modulus, $1/E_{[100]}$ and $1/E_{[111]}$, are 3.881 and 0.475×10^{-12} at room temperature and 4.970 and 0.583×10^{-12} cm²/dyne at the critical temperature (468°C). The elastic anisotropy, given by $E_{[111]}/E_{[100]}$, has the value, 8.2, at room temperature, decreases to a minimum, 7.2, at 300°C, and again increases to 8.6 at the critical temperature. Plots of the reciprocal of Young's modulus in three principal directions, [100], [110], and [111], show that the effect of disorder is mainly felt in the first two directions, while in the third direction the elasticity shows an approximately linear increase with increasing temperature. Before disorder sets in the elasticity in the [100] direction decreases with increasing temperature, contrary to the usual behavior of elastic materials.

INTRODUCTION

DURING the past few years experiments on the change in physical properties of certain ordered alloys as they go through the order-disorder transition have been very fruitful, and the results are in close agreement with theoretical developments in that field. Comprehensive reviews of both theory and experiment have been given by Nix and Shockley,[1] Borelius,[2] and Schottky.[3] While most of the early work was done with polycrystalline material, studies on single crystal specimens have been made recently. Webb[4] has reported a detailed study of the resistivity and thermoelectric power of beta-brass both as a function of temperature and composition. Siegel[5] has obtained the temperature dependence of the principal elastic moduli of Cu_3Au from measurements made on single crystals of the material.

On the theoretical side,[6] no complete treatment has been given of the dependence of elastic properties on the order-disorder transition. It is experimentally established[5,7] that some connection exists, and it is hoped that the present research may throw some light upon the exact nature of this connection and stimulate further theoretical developments.

One of the simplest alloys showing an order-disorder transition is beta-brass. Its crystal structure is body-centered cubic. In the ordered state, zinc atoms occupy cube corners and copper atoms, cube centers, or vice versa. In the disordered state either kind of atom has an equal probability of being in either lattice position. This alloy was chosen for study because Webb's[4] work indicated that a further study of the elasticity of single crystals of this material would be interesting. He found beta-brass to be elastically highly anisotropic in that Young's modulus varied markedly with direction in the crystal lattice. In the present investigation, the anisotropic character of this material was studied more closely at room temperature. Further, the effect of disorder upon Young's modulus for various directions in the crystal lattice was studied.

PREPARATION OF CRYSTALS

The material from which the crystals were grown was supplied through the kindness of Dr. Cyril S. Smith, research metallurgist for the American Brass Company, Waterbury, Connecticut. The accompanying chemical analysis showed the following weight percents: Cu, 51.78; Pb, 0.02; Fe, 0.01; Zn, 48.29 (by difference).

[1] F. Nix and W. Shockley, Rev. Mod. Phys. **10**, 1 (1938).
[2] G. Borelius, Zeits. f. Elektrochemie **45**, 16 (1938).
[3] W. Schottky, Zeits. f. Elektrochemie **45**, 33 (1938).
[4] W. Webb, Phys. Rev. **55**, 297 (1939).
[5] S. Siegel, Phys. Rev. **57**, 537 (1940).
[6] W. Bragg and E. Williams, Proc. Roy. Soc. **A144**, 340 (1934); H. Bethe, Proc. Roy. Soc. **A150**, 552 (1935); R. Peierls, Proc. Roy. Soc. **A154**, 207 (1936).
[7] W. Köster, discussion of paper by Borelius, reference 2.

Neglecting the small impurities, this corresponds to atomic percents of 52.8 and 47.2 for the copper and zinc, respectively. The crystals were grown by Bridgman's[8] method. Instead of making a preliminary polycrystalline casting, a machined blank of the desired size and shape was prepared. It was then coated with several layers of Insalute cement to make an individual mold. In this way crystals with accurately determined dimensions and smooth surfaces were obtained. The blanks were 3 mm square, about 6 cm long and tapered to a point at one end so as to increase the likelihood of a single crystal starting.

Since single crystals of this material are relatively easy to grow, no detailed study was made of growth conditions. It was found, however, that, in general, slower rates of growth produced crystals having a lower orientation function. The rates of growth varied from about 6 cm/hr. to 30 cm/hr. through a temperature gradient of about 75 deg. C/cm. Some twenty-five crystals were grown, approximately fifty percent of which were found to be satisfactory for further use.

Orientation and Orientation Function

The orientation function, $F(l,m,n)$, upon which Young's modulus in any direction depends is

$$F(l,m,n)=l^2m^2+l^2n^2+m^2n^2, \qquad (1)$$

in which l, m, and n are the direction cosines of the length of the specimen with respect to the crystallographic axes. The orientation was determined as described by Webb.[4] Either three or four reflections were used, which gave three or six partially dependent values of l, m, and n. Values of the orientation function calculated from different sets were found to agree within about one percent. The largest error introduced in the measurement results from the inability to set the crystal in the goniometer so that its length is exactly perpendicular to the incident light beam. By removing the crystal, rotating it end for end, and remeasuring the angles, a considerable amount of this error was eliminated. In every case the two different settings gave angles which agreed within one or two degrees. In all cases the crystals exhibited bright reflections when removed from the mold without further treatment. Practically all the observed reflections were from {110} and {111} planes.

Measurement of Young's Modulus

Young's modulus was measured by a dynamic method which has been used extensively and adequately described.[9] Longitudinal vibrations were set up in the crystal specimen by means of a suitably cut square-cross-section quartz rod cemented to it with Insalute cement. Aluminum leaf electrodes were attached to the quartz, and the composite bar was placed on a holder, consisting of a transite block about $6''\times1''\times\frac{1}{2}''$, upon which were mounted two lava knife edges for supporting the bar close to its nodes of vibration. For temperature measurements, three thermocouples[10] were mounted at equally spaced intervals along the length of the holder. As regards frequency the crystals were matched in the usual way. The frequency standardization was made by means of a piezoelectric clock in a manner described by Quimby,[11] and the frequency measurements are believed to be accurate to one part in 100,000.

The runs above room temperature were made by placing the composite oscillator and its holder in a heavy iron cylinder near the middle of an Alundum-cored furnace. The furnace windings were so arranged that the temperature difference over the length of the specimen did not exceed one or two degrees. Readings were taken with both increasing and decreasing temperature. In most cases the crystal was run back and forth through the critical region two or three times. To eliminate lag between the temperature measured by the thermocouples and the actual temperature of the crystal, successive readings were taken as follows. The furnace current was changed slightly and sufficient time allowed to elapse until a steady state ensued, after which the resonance frequency and temperature were measured. At lower temperatures the rate of

[8] P. Bridgman, Proc. Am. Acad. **60**, 305 (1925).

[9] S. L. Quimby, Phys. Rev. **25**, 558 (1925); J. Zacharias, Phys. Rev. **44**, 116 (1933); L. Balamuth, Phys. Rev. **45**, 715 (1934); F. Rose, Phys. Rev. **49**, 50 (1936).

[10] Carefully calibrated Chromel-Alumel thermocouples were used in conjunction with a Leeds and Northrup type K potentiometer.

[11] S. L. Quimby, Phys. Rev. **39**, 345 (1932). The writer is indebted to Professor Quimby for many of the circuit diagrams used and many helpful suggestions.

heating or cooling was about 25°C per hour. In the neighborhood of the critical temperature the rate was less. Measurements made with increasing and decreasing temperatures agreed so closely that it was certain that the lag was eliminated and that the thermocouples measured the true temperature of the specimen. The relaxation time of beta-brass is known to be so short that no lag due to departure from an equilibrium condition needed to be guarded against.

Young's modulus along the axis of a bar at temperature T, is related to the resonance frequency, f, of the brass specimen by the formula,[5]

$$E=4\rho_{25}L^2{}_{25}f\left(\frac{L_{25}}{L_T}\right)\left[1+\frac{1}{2}\frac{A\bar{\sigma}^2n^2}{L^2{}_{25}}\right], \qquad (2)$$

where ρ_{25} and L_{25} are the density and lengths of the bar at 25°C, respectively, $\bar{\sigma}$ is the average Poisson's ratio, n is the approximate number of half-waves in the metal crystal, and A is the area of the cross section. The term in the brackets is a correction due to the lateral motion. It is so near unity that in the present calculation it has been taken as one. In order to get L_T from L_{25}, values of the coefficient of thermal expansion of polycrystalline beta-brass, determined by Merica and Schad,[12] were used.

A comparator was used to measure the lengths of the specimens at room temperature. Eight measurements were taken on each crystal and the uncertainty in length is believed not to exceed 0.1 percent.

The density at 25°C was determined by the method of hydrostatic weighing. The value, 8.35 g per cc, so obtained is, judging by the deviations of individual readings from the mean, accurate within 0.5 percent.

TABLE I. *Summary of data for seven single crystals at room temperature.*

CRYSTAL	$F(l, m, n)$	$1/E\times10^{12}$ CM²/DYNE	$E\times10^{-11}$ DYNE/CM²
6	0.0220	3.664	2.729
7	0.0780	3.082	3.245
24	0.1648	2.202	4.541
22	0.1868	1.938	5.160
11	0.2439	1.418	7.052
16	0.2665	1.148	8.711
12	0.2756	1.069	9.355

[12] P. Merica and L. Schad, Bull. Bur. Stand. **14**, 571 (1918). The values they give are 22.8, 19.4, 23.5, 27.5, 39.2, and 26.9×10^{-6} per deg. C for the temperature ranges 25–100, 100–200, 200–300, 300–400, 400–469, and 470–510 degrees C, respectively.

RESULTS AT ROOM TEMPERATURE

Table I gives results for room temperature. Similar data were obtained for temperatures ranging from room temperature to above the critical temperature for order.

The principal elastic coefficients,[13] s_{11}, s_{12}, and s_{44} of a cubic crystal are related to Young's modulus by the formula

$$1/E=s_{11}-2sF(l,m,n), \qquad (3)$$

where $s=s_{11}-s_{12}-\frac{1}{2}s_{44}$

and $F(l,m,n)$ is the orientation function previously given. Equation (3) shows that a plot of $1/E$ against $F(l,m,n)$ should be a straight line. From it, one of the elastic coefficients, s_{11}, and a relationship between the other two, namely, $(2s_{12}+s_{44})$, may be determined. Quantities of the most interest physically, perhaps, are the reciprocals of Young's modulus in the [100], [110], and [111] directions. The first of these is the elastic coefficient, s_{11}, and corresponds to a direction along the cube edge of the crystal. The second corresponds, among other directions, to the direction along a face diagonal, and the third to a direction along the cube diagonal. Equation (3) yields $1/E_{[111]}$, $1/E_{[110]}$, and $1/E_{[100]}$ $(=s_{11})$ upon substitution of $\frac{1}{3}$, $\frac{1}{4}$, and zero, respectively, for the orientation function.

The reciprocal of Young's modulus of each of the seven crystals was plotted against its orientation function for about thirty different temperatures. At each temperature, a straight line was fitted to the points by the method of least squares. For room temperature, the appropriate values from Table I are shown plotted as curve *B*, Fig. 1.

Insofar as the writer knowns, the only other data available on the elastic properties of beta-brass single crystals are those of Webb.[4] He states that his determination can hardly be con-

[13] Schmid and Boas, *Struktur und Eigenschaften der Materie XVII, Kristallplastizität*, p. 23. These s_{11}, s_{12}, etc. are variously known as elastic constants, moduli, coefficients, parameters, etc. The writer will adopt the terminology used by Schmid and Boas and call them coefficients. The term "modulus" does not seem so happy in that, dimensionally, the s's are strain/stress, whereas modulus is commonly used for stress/strain.

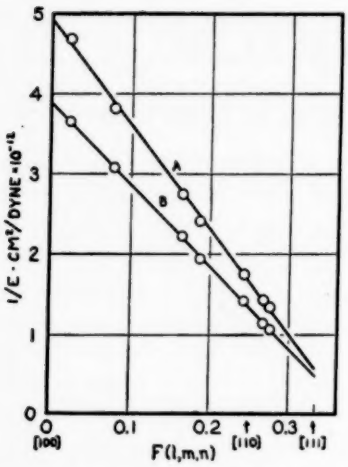


FIG. 1. $1/E$ against orientation function. Curve B at room temperature. Curve A at 466°C, temperature at which anisotropy in Young's modulus is a maximum.

sidered complete or of very high precision chiefly because of the irregularity of his specimens. In the final fitting of the curve, Webb ignored two of his points solely because they did not seem to agree with the rest of the observations. To the writer, it seems preferable to include these points, and a least-square fit has been made on this basis. Table II compares Webb's values as given by him and as corrected, with the writer's.[14] The corrected values are seen to be in better agreement with the writer's.

The unusual degree of anisotropy in Young's modulus, as shown by the large ratio of maximum to minimum modulus in the [111] and [100] directions, respectively, was noted by Webb.[4] The writer's results check Webb's reasonably well. Webb suggested further, that this high degree of anisotropy might disappear in the disordered state.[15] The writer's results provide a test of this hypothesis. It was found to be incorrect as will appear later.

DEPENDENCE ON TEMPERATURE

The results of the measurements at elevated temperatures are presented in summarized form in Table III and shown graphically in the curves of Figs. 2 and 3, which are plots of reciprocals of Young's modulus in three directions in the crystal, [100], [110], and [111]. The behavior of these three curves is most conveniently described by dividing the temperature range into the four rather arbitrary regions, 25°C–150°C, 150°C–468°C, 468°C–469°C, and 470°C–510°C.

In the first of these regions, specific heat[16] and resistivity[4] measurements indicate that very little, if any, disordering sets in. It might be expected, then, that any change in elasticity would be the linear increase in s_{11}, s_{12}, etc. observed in most substances. It is apparent from the curves, that, although the part below 150°C is approximately straight, the slope depends entirely upon the orientation of the crystal. In the [111] direction, the behavior of Young's modulus is like that for an isotropic solid (Fig. 3). In the [110] direction and other directions with the same orientation function (see lower curve in Fig. 2) there is almost zero change in $1/E$. Qualitatively, this is in agreement with Köster,[7] who found very little change in Young's modulus with temperature over this temperature range for polycrystalline beta-brass, since a specimen consisting of randomly oriented crystals would

[14] It must be remembered, however, that the present method is a dynamic one and it is the adiabatic Young's modulus which is obtained, while the static method used by Webb gives the isothermal value. The difference between the two is easily calculated (see Voigt, *Lehrbuch der Kristallphysik*, p. 789) and is found to be much too small to account for the difference between Webb's and the writer's values. Another reason for the discrepancy may lie in the fact that Webb's crystals varied widely in composition, but neither the writer nor Webb was able to find from his data any direct correlation between the composition and the value of Young's modulus.

[15] Webb's argument was that, while the large anisotropy at room temperature might easily be attributed to a difference in binding energy between like atoms and unlike atoms, in the disordered state, the various kinds of possible bindings might be expected to give a sort of average result, and hence make the crystal more nearly isotropic.

[16] H. Moser, Physik. Zeits. **37**, 737 (1936).

TABLE II. *Comparison with Webb's results.*

	$1/E_{[100]}\times10^{12}$ CM²/DYNE	$1/E_{[110]}\times10^{12}$ CM²/DYNE	$1/E_{[111]}\times10^{12}$ CM²/DYNE	$(2s_{12}+s_{44})\times10^{12}$ CM²/DYNE	$E_{[111]}/E_{[100]}$
Webb	4.48	1.53	0.50	−2.8	8.9
Webb (corrected)	4.28	1.48	0.54	−2.7	7.9
Rinehart	3.88	1.33	0.48	−2.5	8.2

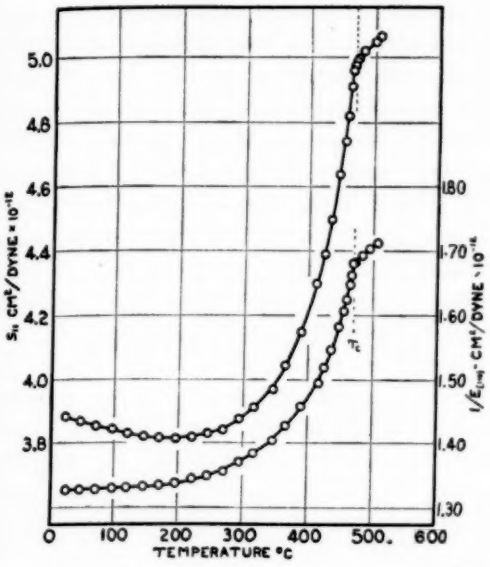


FIG. 2. $1/E$ against temperature. Lower curve, $1/E_{[110]}$. Upper curve, $1/E_{[100]}$.

TABLE III. *Reciprocal of Young's modulus in principal directions.*

TEMP. °C	$1/E_{[100]} \times 10^{12}$ CM²/DYNE	$1/E_{[110]} \times 10^{12}$ CM²/DYNE	$1/E_{[111]} \times 10^{12}$ CM²/DYNE
24	3.881	1.326	0.475
48	3.865	1.326	.480
72	3.850	1.327	.486
96	3.841	1.327	.489
121	3.829	1.328	.495
145	3.817	1.330	.502
170	3.814	1.333	.506
195	3.810	1.337	.512
220	3.815	1.342	.517
245	3.826	1.348	.522
269	3.838	1.354	.526
293	3.871	1.368	.534
317	3.909	1.383	.541
341	3.967	1.402	.546
365	4.042	1.426	.554
389	4.148	1.457	.560
412	4.296	1.495	.561
424	4.386	1.519	.563
436	4.492	1.547	.565
448	4.637	1.581	.562
455	4.738	1.606	.561
459	4.816	1.623	.559
464	4.908	1.647	.560
466	4.959	1.662	.563
468(T_c)	4.970	1.680	.583
471	4.987	1.682	.580
473	4.993	1.683	.579
483	5.020	1.692	.583
495	5.045	1.702	.588
506	5.064	1.710	.593

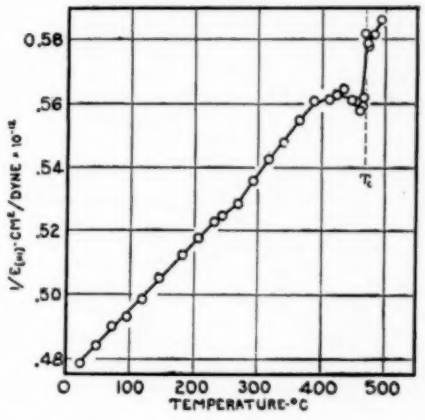


FIG. 3. $1/E_{[111]}$ against temperature.

necessarily contain a large proportion of crystals with orientation function near 0.250. The reciprocal of Young's modulus in the [100] direction, shown in the upper curve of Fig. 2, decreases rapidly. This is very surprising in view of the fact that in almost every other known material just the reverse is true. Insofar as the writer knows, no theories as yet have been proposed which would predict such an abnormal behavior as this. It seems not unlikely that a close connection may exist between the unusual anisotropy exhibited by beta-brass and the peculiar temperature dependence of the elasticity in this temperature region. The different behaviors of these three curves near room temperature suggest that, perhaps, measurements on the material at much lower temperatures might be of some interest. Such an investigation is planned.

Above 150°C, the very rapid increase in slope in $1/E_{[100]}$ $(=s_{11})$ and $1/E_{[110]}$, Fig. 2, is undoubtedly due to the disappearance of order, and is in close agreement with the experimental results of Siegel[5] for Cu_3Au and Köster[7] for polycrystalline beta-brass. Köster found that disordering caused the value of $1/E$ to be about 30 percent higher at 470°C than its value at 150°C, and the writer's results indicate that over the same temperature interval, $1/E_{[110]}$ increases about 37 percent. The behavior of $1/E_{[111]}$, as shown in Fig. 3, is quite different from the behavior of either of the curves in Fig. 2. The linear increase, observed between 25°C and 150°C, is extended right up to 400°C; therefore, disordering presumably produces no particular

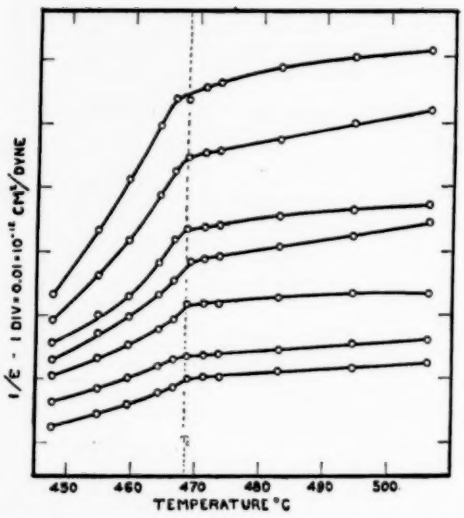


FIG. 4. $1/E$ against temperature for individual crystals in the neighborhood of the critical temperature. Values at critical temperature are, reading from top to bottom (increasing orientation function): 4.686, 3.846, 2.758, 2.433, 1.791, 1.433, and 1.349×10^{-12} cm²/dyne, respectively.

effect in this region. Between 400°C and the critical temperature, 468°C, the curve shows a decided dip with a subsequent sharp rise at the critical temperature. It may be that this dip in the curve is only a result of the computational method used; if, however, it is real, only the last sudden disappearance of order modifies the steady rise of this quantity. In any case the effect of disordering is far less noticeable in this direction than it is in any other direction in the crystal. No crystal oriented exactly in the [111] direction was available, so that the reality of the dip could not be confirmed by a direct, independent run on a crystal.

An abrupt change in slope for each of the three curves in Figs. 2 and 3 occurs between 468°C and 469°C. This is due to the final disappearance of order. There is evidently no discontinuity in the curve such as has been observed in similar curves for Cu_3Au.[5] This is in agreement with the fact that discontinuities in specific heat[17] and resistivity[5] are found for Cu_3Au but not for beta-brass. Webb[4] points out the existence of a fairly well defined "tail" on his resistivity *vs.* temperature curves. In Fig. 4, the reciprocal of Young's modulus against temperature is plotted, on a greatly enlarged scale, in the neighborhood of the critical temperature for each of the seven crystals run. There is no evidence of a tail of width greater than two degrees, which is the maximum difference in temperature between the two ends of the crystal. It is very likely, therefore, that the slope is discontinuous at the critical temperature.

Above the critical temperature, the reciprocal modulus shows the normal behavior, i.e., it increases with temperature for any direction in the crystal lattice. (See curves in Figs. 2, 3, and 4.)

The dependence of critical temperature upon relative concentration of zinc and copper has been measured by Sykes and Wilkinson.[18] They find the critical temperature for the composition used in this investigation to be 468°C. This agrees closely with the value obtained by the writer, which certainly lies somewhere between 468°C and 469°C.

ANISOTROPY AND TEMPERATURE

While the degree of anisotropy of the crystal varies with temperature, as can be seen from the results already cited, it is of some interest to consider it specifically. Taking as before, $E_{[111]}/E_{[100]}$ as a measure of the anisotropy, its variation with temperature is shown, graphically, in Fig. 5. While initially the anisotropy does decrease slightly, it increases again rapidly in the very region where the disordering is proceeding most rapidly. Thus, it appears that the simple argument advanced to support Webb's[15]

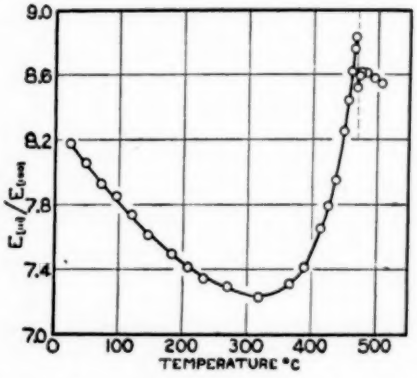


FIG. 5. Anisotropy in Young's modulus against temperature.

[17] The discontinuity in specific heat is expected on theoretical grounds, reference 1, Section 14.

[18] C. Sykes and H. Wilkinson, J. Inst. Metals **61**, 223 (1937).

hypothesis is entirely too simple a picture. The rapid fluctuation in the curve just at the critical temperature is due to the previously discussed dip in the curve of Fig. 3. The same remarks as to its genuineness would apply. Just beyond the critical temperature, the anisotropy begins to decrease with increasing temperature just as it does in the neighborhood of room temperature. A qualitative picture of the temperature dependence of this anisotropy can be obtained by considering the behavior of the lower line of Fig. 1, as temperature is increased. The curve rotates counterclockwise about the value of $F(l,m,n)$, which corresponds to the [110] direction in the crystal lattice, until the temperature at which disorder begins to set in is reached, i.e., 150°C. As disordering proceeds, the rotation becomes clockwise and, in addition, the whole curve shifts upward until finally at the critical temperature it has risen considerably above its position at room temperature and possesses a much greater slope. Curve A, in Fig. 1, is a plot at a temperature just below the critical temperature, where the anisotropy is a maximum.

In conclusion, the writer gratefully acknowledges his indebtedness to Professor E. P. T. Tyndall for much helpful advice and many valuable suggestions throughout the work; to the Physics Department of the State University of Iowa for the facilities generously placed at his disposal; and to others who assisted him in this work.

AUGUST 15, 1940 PHYSICAL REVIEW VOLUME 58

The Internal Friction of Single Metal Crystals*

THOMAS A. READ
Pupin Physics Laboratories, Columbia University, New York, New York

(Received June 1, 1940)

The internal friction of crystalline copper, tin, lead, and zinc has been measured by the composite piezoelectric oscillator method. It is found that the decrement of an unannealed crystal may be as large as that of the polycrystalline material, that annealing reduces the decrement to a value of the order 10^{-4} to 10^{-5}, and that both Young's modulus and the decrement vary with the vibrational strain amplitude at strain amplitudes as low as 10^{-6}. In the case of zinc crystals, a detailed study has been made of the way in which the elastic modulus and internal friction depend on the previous history of the specimen, on the vibration frequency and amplitude, and on the orientation of the vibration axis with respect to the crystal slip planes. The results suggest that the mechanism involved is a propagated "dislocation" of the sort proposed by Taylor, Polanyi and Orowan to account for macroscopic plastic flow, and that the application of a stress is accompanied by a plastic strain, together with an associated strain hardening in consequence of which the stress-strain relation on removal of the applied stress is nearly elastic.

INTRODUCTION

INTERNAL friction is a property of matter in virtue of which organized energy of imperfectly elastic stress is rendered irrecoverable, or unavailable for mechanical work. Thus a specimen body which is carried adiabatically through a complete stress cycle can be restored to its initial state only by the removal of a certain amount of heat, and this heat, called the "energy dissipated per cycle," affords a measure of the internal friction.

The "coefficient of internal friction," ξ, of a substance is defined by the formula[1]

$$W^d = \xi S^2, \tag{1}$$

where W^d is the energy dissipated per cycle per unit volume, and S is the stress amplitude. The internal friction of solid materials is usually studied by observing the behavior of a properly

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

[1] A. L. Kimball and D. E. Lovell, Phys. Rev. **30**, 948 (1927).

shaped specimen body which is vibrating in one of its normal modes. Such experiments show that the value of ξ depends on the type of stress; on the frequency, and possibly the amplitude, of vibration; and, in ferromagnetic substances, on the magnetization.[2]

Considerable progress has been made toward an understanding of the processes responsible for the dissipation of vibrational energy in solids. Thus an adequate account has been given of the magnetization-dependent part of the internal friction in ferromagnetic materials,[3] and of the greater part of the internal friction in nonferromagnetic polycrystalline metals. The latter is largely the work of Zener and his collaborators.[4] According to Zener's theory, microscopic stress inhomogeneities due to occluded impurities, cold working, or simply microcrystalline anisotropy, give rise to local temperature fluctuations in the vibrating material, and hence to heat flow, with a consequent irreversibility in the stress-strain relation. However, the experiments reveal that this is not the only dissipative mechanism operative in the medium, but that a supplementary process must be invoked to account for a residual internal friction which, though small, is still many times larger than that of crystalline quartz.[5] The present study of the internal friction in large single crystals of metals contains a clue to the nature of this process. For it is found that both plastic flow and strain hardening occur in these crystals at vibration frequencies as high as 70 kilocycles per second, and vibrational strain amplitudes as low as 10^{-6} cm/cm.

[2] *Normal:* A. L. Kimball and D. E. Lovell, Phys. Rev. **30**, 948 (1927); R. L. Wegel and H. Walther, Physics **6**, 141 (1935); F. Forster and W. Koster, Zeits. f. Metallkunde **29**, 116 (1937), Naturwiss. **25**, 436 (1937). *Ferromagnetic:* In iron, W. T. Cooke and W. F. Brown, Phys. Rev. **50**, 1158 and 1165 (1936); in nickel, S. Siegel and S. L. Quimby, Phys. Rev. **49**, 663 (1936); in crystalline nickel, J. R. Zacharias, Phys. Rev. **44**, 116 (1933); in permalloy, S. Siegel and S. Rosin, Phys. Rev. **49**, 863 (1936). *Of crystalline quartz:* K. S. Van Dyke, Proc. I. R. E. **23**, 386 (1935). *During the lambda-point transition in ammonium chloride;* A. W. Lawson, Phys. Rev. **57**, 417 (1940).

[3] W. F. Brown, reference 2.

[4] C. Zener, Phys. Rev. **52**, 230 (1937); **53**, 90 (1938); **53**, 100 (1938); **53**, 582 (1938); **53**, 1010 (1938); Am. Inst. Mining and Metallurgical Eng., Technical Publication No. 1146.

[5] K. S. Van Dyke, reference 2.

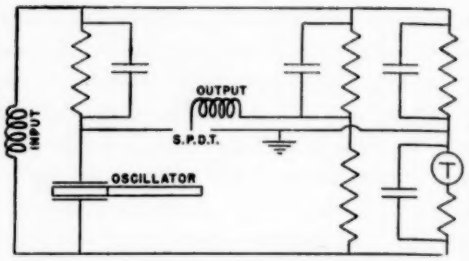


FIG. 1. Diagram of the alternating-current bridge, showing the vacuum thermocouple, T, in one arm of the Wagner ground connection.

EXPERIMENTAL METHOD

It has been remarked that the value of the internal friction coefficient depends in the first instance upon the nature of the stress to which the medium is subjected. The type of stress here dealt with is the stretch associated with the longitudinal vibration of a slender rod of specimen material, and the experimental method permits the measurement of both ξ and the Young's modulus, E. This method is essentially that used by Cooke and Brown,[2] so modified as to permit the investigation of nonhomogeneous specimens. The nonhomogeneity results from the fact that, in the materials here studied, the values of both ξ and E *depend on the strain amplitude*, and the latter varies sinusoidally along the vibrating rod. The method applicable to homogeneous materials will be described first, and then its extension to meet the demands of the present problem.

Homogeneous specimen

The specimen, in the form of a right circular cylinder 5.38 mm in diameter and a few cm long, forms one part of a composite piezoelectric oscillator constructed by cementing to one end of the specimen an X cut cylinder of crystalline quartz of identical cross section. The oscillator forms one arm of an alternating-current bridge (Fig. 1), which is excited by a very stable vacuum tube oscillator whose frequency is variable between 30 kc and 80 kc. The electric impedance of the composite oscillator varies critically with frequency in the neighborhood of certain "resonance frequencies," and the elastic and dissipative properties of the specimen material are deduced from the observed nature of this variation.

It can be shown that, near a resonance frequency, the composite oscillator is electrically equivalent to a fixed capacity shunted by a series resonant electric circuit. If the vector impedance of the latter be denoted by $R+iX$, then the resonance frequencies, f_0, are, by definition, those for which $X=0$, and these are the solutions for f_0 of the equation

$$M_1 f_1 \tan(\pi f_0/f_1)+M_2 f_2 \tan(\pi f_0/f_2)=0, \quad (2)$$

where

$$\begin{aligned} M_i &= \text{mass of a cylinder,} \\ f_i &= (E_i/\rho_i)^{\frac{1}{2}}(2L_i)^{-1}, \\ E_i &= \text{Young's modulus,} \\ \rho_i &= \text{density of a cylinder,} \\ L_i &= \text{length of a cylinder,} \end{aligned} \quad (3)$$

and the subscripts 1 and 2 refer to the specimen and quartz cylinders, respectively. In consequence of the piezoelectric stress which accompanies the electric field in the quartz, a stationary state of forced longitudinal vibration is established in the composite oscillator, and the amplitude of this elastic vibration is a maximum at a resonance frequency.

The inductance, $\mathfrak{L}$, and the capacity, C, of the equivalent series resonant circuit are given by the formulae

$$\mathfrak{L}=\tfrac{1}{8}K(M_1+M_2) \quad (4)$$

and

$$C=2/[\pi^2 f_0^2 K(M_1+M_2)], \quad (5)$$

where K is a quantity whose value depends only on the appropriate piezoelectric constant of the quartz and the geometry of the electrodes.[6] The resistance, R, of this circuit, which is also the resistance of the composite oscillator as measured with the bridge, is conveniently evaluated with the aid of two auxiliary quantities, Δ_1 and Δ_2, associated with the specimen and quartz cylinders respectively and defined by the formula

$$\Delta_i = W_i^d/2W_i^v, \quad (6)$$

where W_i^d denotes the energy dissipated per cycle, and W_i^v the total vibrational energy, in a cylinder. The Δ_i's are called the "decrements" of the cylinders, and it can easily be shown that if, and only if, ξ is independent of the stress amplitude then

$$\Delta_i=\xi_i E_i, \quad (7)$$

and hence that, in this event, Δ_1 is a measure of a property both of the specimen and the specimen material. The resistance is given by the formula

$$R=\tfrac{1}{4}K(M_1\Delta_1+M_2\Delta_2)f_0, \quad (8)$$

and it follows from Eqs. (4) and (8) that

$$R/2f_0\mathfrak{L}=(M_1\Delta_1+M_2\Delta_2)/(M_1+M_2). \quad (9)$$

It will be noted that the left-hand member of this equation is, by definition, the electric decrement, Δ_e, of the equivalent series resonant circuit, and so is a quantity which, like R, is directly measureable with the bridge. Thus

$$\Delta_e=R/2f_0\mathfrak{L}, \quad (10)$$

and the value of $\mathfrak{L}$, if desired, follows immediately from measurements of R and Δ_e.

The foregoing formulae, with M_1 set equal to zero, describe the behavior of the quartz cylinder alone, without the specimen attached. Thus, by Eqs. (2), (9) and (10), f_2 is the frequency corresponding to zero reactance, and Δ_2 the decrement, of the single quartz crystal oscillator. When these have been measured, the only additional data necessary to obtain a complete description of the behavior of the composite oscillator are f_0 and Δ_e. For f_1 then follows from Eq. (2), E_1 from (3), Δ_1 from (9) and (10), and ξ_1 from (7).

Inhomogeneous specimen

It has been remarked that the inhomogeneity of the present specimens arises from the fact that the values of both ξ and E depend upon the local vibrational strain amplitude, which, in turn, varies sinusoidally along the rod. The following treatment is based upon two important considerations: first, that while, over the hundred-fold range of variation of strain amplitude here explored, the variation of ξ may be a hundred-fold, that of E is only of the order 0.04 percent; and second, that the lengths of the quartz and specimen rods are so adjusted that at resonance there are, with a discrepancy not greater than one percent, integral numbers of half-waves of vibra-

[6] $K=1/(b^2\epsilon^2)$, in an example with idealized geometry, where ϵ is the piezoelectric constant and b is the width of an electrode on the quartz. The value of K for the present experimental arrangement is 10,300 at 38 kc, where R is in ohms, $\mathfrak{L}$ in henries, and C in farads.

tion in the quartz and specimen rods. In other words, under the conditions of measurement both the quartz and specimen cylinders are vibrating almost exactly in one of their respective normal modes.

The first question which will be considered concerns the measurement of the equivalent ioductance $\mathfrak{L}$. When the specimen is homogenenus, $\mathfrak{L}$ is calculated with Eq. (10), from the measured resistance at resonance and electric decrement. The latter quantity is evaluated by observing the frequency departure from resonance, δf, at which the scalar impedance is increased in the ratio $\sqrt{2}:1$. Then $\Delta_e = 2\pi\delta f/f_0$, but this relation no longer holds when Δ_1, and hence R, depends on the amplitude of vibration. Fortunately the materials here studied are of such nature that it is possible to find a range of vibration amplitudes sufficiently small that the variation of ξ with amplitude over the range is negligible. It remains only to note, (Eq. (4)), that $\mathfrak{L}$ depends on the properties of the specimen only through the mass, and hence that a single valid measurement of $\mathfrak{L}$ suffices at all amplitudes. This conclusion has been verified experimentally.

The second question concerns the relation of ξ_1 to Δ_1, and of Δ_1 to R. Δ_1 remains a property of the specimen, but is no longer a property of the material. However, a quantity, Δ, which is a property of the material can be defined by the formula

$$\Delta = \xi E. \tag{11}$$

It then follows from Eqs. (1) and (6) that

$$\Delta_1 = \int_0^{L_1} \Delta U_1^2 dx \div \int_0^{L_1} U_1^2 dx, \tag{12}$$

where $U_1 (= S_1/E_1)$ denotes the local strain amplitude, dx an element of length on the cylinder axis, and the vibrational energy density in the specimen has been equated to $\frac{1}{2}E_1U_1^2$. In accordance with prevalent practice, Δ rather than ξ will hereafter be used to describe the dissipative property of a substance.

The resistance of the oscillator is given by the formula (8), whether the specimen is homogeneous or not. This assertion can best be understood by observing that energy dissipation of whatever sort which occurs in the specimen, is manifested only through a dissipative stress acting on the quartz at the quartz-specimen interface. By dissipative stress is meant one in phase with the particle velocity. Accordingly, the question may be phrased thus: If one end of a quartz oscillator be subjected to a stress which dissipates energy at the rate (Eq. (6)) $W_1^d = 2\Delta_1 W_1^v$ per cycle, will the resistance of the oscillator be increased by the amount $\frac{1}{4}KM_1\Delta_1$ in consequence of this stress? The affirmative result then follows by direct calculation.

The third question concerns the meaning of the quantity f_1 which appears in Eq. (2), and its relation to the elasticity of the specimen. Here f_1 is defined as the solution of Eq. (2) which corresponds to a given (observed) value of f_0, and is also, by Eq. (3), the fundamental frequency of free longitudinal vibration of a *homogeneous* specimen. It will now be shown that, under the aforementioned conditions of measurement, f_1 is very nearly the fundamental frequency of free vibration of an elastically inhomogeneous specimen.

Equation (2) is usually obtained by a straightforward mathematical development in which the behavior of the composite oscillator is regarded as a problem in the theory of elasticity. The same result may be derived more easily from the following significant observation concerning the factors which determine the resonant frequency of a composite oscillator, *viz.*: The fundamental resonant frequency of a composite oscillator composed of lengths L_1 of homogeneous specimen material and L_2 of quartz is the common fundamental resonant frequency of lengths L_1^0 of specimen material and L_2^0 of quartz so chosen that, when both rods are vibrating in their respective fundamental modes, the stress and displacement amplitudes at distances L_1 from one end of L_1^0 and L_2 from one end of L_2^0 are respectively equal to each other and to the stress and displacement amplitude at the quartz-specimen interface of the composite oscillator. For under these circumstances lengths L_1 and L_2 of the rods, resonating at the common frequency f_0, can be imagined to be cut off and cemented together without disturbing the vibrational state, since the boundary conditions of continuity of stress and displacement at the interface are met.

The continuity of stress at the interface yields the equation

$$(E_1\pi u_1^0/L_1)\sin(\pi L_1/L_1^0) + (E_2\pi u_2^0/L_2)\sin(\pi L_2/L_2^0)=0, \quad (13)$$

and the continuity of displacement the equation

$$u_1^0\cos(\pi L_1/L_1^0)=u_2^0\cos(\pi L_2/L_2^0), \quad (14)$$

in which u_i^0 denotes the amplitude of vibration of a rod. Accordingly

$$(E_1/L_1^0)\tan(\pi L_1/L_1^0) + (E_2/L_2^0)\tan(\pi L_2/L_2^0)=0. \quad (15)$$

Now $E_i=\rho_i V_i^2$, where $V_i=2f_0L_i^0=2f_iL_i$, hence

$$f_0/f_i=L_i/L_i^0, \quad (16)$$

and Eq. (2) follows immediately from Eq. (15). This form of development is particularly suited to reveal the effect of specimen inhomogeneity on the significance of f_1.

It has been remarked before that the type of inhomogeneity encountered in the present research is a small variation of E_1 with the local strain amplitude, and also that, under the conditions of measurement, the specimen is vibrating very nearly in a normal mode. Thus the space variation of the particle displacement near the interface, where the strain is nearly zero, is practically unaltered by the inhomogeneity and Eq. (15) holds, with E_1 interpreted as the Young's modulus for zero strain. Now it is shown in the next paragraph, and is also intuitionally evident, that the resonant frequency of a rod depends chiefly on its elasticity at the displacement *nodes* of vibration and the mass per unit length at the displacement *loops*. Hence Eq. (16) likewise holds very nearly, with f_0 and f_1 interpreted as the resonant frequencies of inhomogeneous rods of lengths L_1^0 and L_1, respectively.[7] However, E_1 can no longer be set equal to $4\rho_1 f_1^2 L_1^2$ exactly, and consequently the value of M_1 in Eq. (2) is uncertain to the extent of the entire variation in E_1, i.e. about 0.04 percent. Nevertheless, the experimental conditions that $L_1=L_1^0$ and $L_2=L_2^0$ very nearly imply that the resulting error in the calculated value of f_1 is negligible. For under these conditions

$$\tan(\pi f_0/f_1)=\pi(f_0-f_1)/f_1,$$
$$\tan(\pi f_0/f_2)=\pi(f_0-f_2)/f_2,$$

and f_1 is given, from Eq. (2), by the formula

$$f_1=f_0+(M_2/M_1)(f_0-f_2). \quad (17)$$

In a typical example, $f_0=40$ kilocycles and (f_0-f_2) 200 cycles. Hence an uncertainty of 0.04 percent in M_1 produces an uncertainty of 0.08 cycles in the *absolute* value of f_1, and an uncertainty of 0.04 percent in the calculated value of the *change* in f_1 associated with a change in the amplitude of vibration. It is only the latter quantity that is of special interest in the present investigation.

It remains to relate f_1 to the elasticity of the inhomogeneous rod. This is accomplished by applying the usual approximation method[8] to integrate the differential equation $\partial(E\partial u/\partial x)/\partial x=\rho\partial^2 u/\partial t^2$, with the result that

$$\delta f_1/f_1=(1/L_1)\int_0^{L_1}(\delta E_1/E_1)\sin^2(\pi x/L_1)dx, \quad (18)$$

where f_1 is the resonant frequency of a homogeneous rod of elasticity E_1, and δf_1 is the change in f_1 associated with the change δE_1 in E_1.

Measurement of the maximum strain amplitude

The calculation of the maximum strain amplitude in the specimen is based on the observation that the electric decrement of the equivalent electric circuit, as measured on the bridge, is equal to the ratio of the energy dissipated per cycle in the composite oscillator, W_0^d, to twice the total vibrational energy at resonance, W_0^v. Then

$$W_0^d=10^7\mathcal{E}^2/Rf_0=2\Delta_e W_0^v=RW_0^v/f_0\mathfrak{L},$$

where $\mathcal{E}$ denotes the root-mean-squared voltage applied to the quartz. W^v_0 is readily evaluated in terms of the maximum strain amplitude in the specimen, U_1^0, and it follows that

$$U_1^0=6.325\times10^3(\mathcal{E}/R)(\mathfrak{L}/V_1E_1)^{\frac{1}{2}} \times[1+(m/n)(E_2\rho_2/E_1\rho_1)^{\frac{1}{2}}]^{-\frac{1}{2}}, \quad (19)$$

where m and n denote the number of half-waves

[7] The error is of the order $(\delta f/f)^2$, where δf is the change in resonant frequency due to the inhomogeneity.

[8] See, for example, J. C. Slater, *Introduction to Theoretical Physics* (McGraw-Hill, 1933), p. 154.

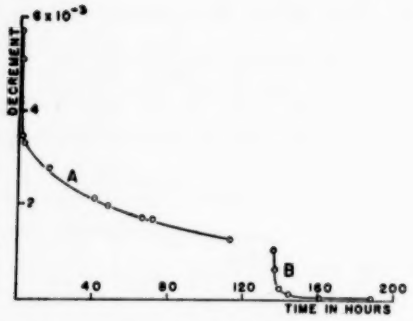


FIG. 2. Showing the change in the decrement of an isolated zinc crystal with time; Part A, before annealing; Part B, after annealing.

of vibration in the quartz and specimen, respectively.

Experimental details

The quartz is etched in the manner described by Van Dyke[5] to minimize Δ_2, and cemented to the specimen with a very thin film of beeswax and rosin. The composite oscillator rests horizontally on two fine silk threads placed at displacement nodes of vibration, with the quartz crystal between, but not touching, parallel plate electrodes. The environment of the oscillator is evacuated and temperature stabilized to 0.01°C at 26°C.

The frequency of the voltage applied to the oscillator is measured by comparison with the output of a piezoelectric clock.[9] The absolute value of the frequency is known only to one part in 10^5, but frequency *changes* can be measured with an accuracy better than 0.01 cycle.

The voltage, $\mathcal{E}$, applied to the quartz is measured with the aid of a calibrated vacuum thermocouple placed in one arm of the Wagner ground connection, as shown in Fig. 1. The bridge detector is sufficiently sensitive to permit the use of the bridge when $\mathcal{E}$ is as small as 0.001 volt.

SPECIMEN MATERIAL

The copper crystals are grown from commercial copper wire. The analyses of the remaining materials, as supplied by the makers, are as follows: tin (Eimer and Amend Co.)—percent impurity, Fe 0.004, Zn 0.004, Pb 0.001, As 0.00001; lead (Eimer and Amend Co. "test lead") silver, gold and bismuth free; zinc (New Jersey Zinc Co.)—percent impurity, Fe 0.0004 to 0.0009, Pb less than 0.0002, Cd less than 0.00005.

The copper crystals are grown in a graphite crucible, in the furnace described by Quimby.[9]

The tin, lead and zinc crystals are grown by Miller's modification of Bridgman's method.[10] The principal features of this method are the following: The material is melted in a Pyrex glass flask, and thoroughly degassed in a vacuum. It is then cast in cylindrical molds, made of Pyrex glass tubing, which have been thoroughly cleaned and coated inside with colloidal graphite to prevent the material from sticking to the walls. The loaded molds are packed in Sil-O-Cel, in a large test tube, and lowered through an electric furnace at the rate of a quarter of an inch per hour.

RESULTS

The internal friction of well-annealed single crystals is considerably less than that of the corresponding polycrystalline materials as reported by other observers.[2] Thus the values of Δ_1 at small amplitudes of vibration are, for copper—3.6×10^{-5}; for tin—6.9×10^{-5}; for lead —2.8×10^{-4}; and for zinc—1.0×10^{-5} to 2.2×10^{-5}.[11] The decrements of unannealed single crystals may be as large as those of polycrystalline specimens. For example, the decrement of a freshly prepared copper crystal may decrease from 3×10^{-3} to the value given above, i.e., almost a hundred-fold, after it is annealed for two hours in hydrogen at 500°C.

But the most significant dissipative characteristic of these crystals is an associated variation of the decrement and elastic modulus with the vibrational strain amplitude. It will later be suggested that this phenomenon is due to

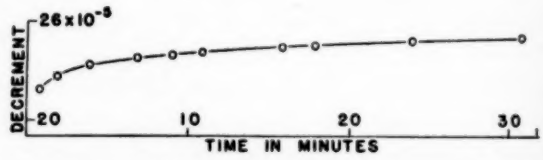


FIG. 3. Showing the change in the decrement of a continuously oscillating zinc crystal with time.

[9] See S. L. Quimby, Phys. Rev. **39**, 345 (1932).

[10] P. W. Bridgman, Proc. Am. Acad. **60**, 305 (1925); R. F. Miller, Trans. Inst. of Metals Div. Am. Inst. Min. Eng. **122**, 176 (1936).

[11] A preliminary report of these measurements was given in Phys. Rev. **54**, 389 (1938).

sporadic displacements of atoms or groups of atoms parallel to the slip planes of the crystal. If this hypothesis is correct, the effect should be closely correlated with the relative orientation of the vibration axis and the slip planes. Now the crystalline structure of zinc is hexagonal, with a single slip plane perpendicular to the hexagonal axis. Accordingly this substance is particularly suited for the present investigation, and the remainder of this research is concerned with a study of the phenomenon as it appears in single crystals of zinc.

Definitive observations are not easily made, for the reason that almost anything that is done to the crystal leaves an almost indelible imprint on its subsequent behavior. The nature of the difficulties encountered is illustrated by the curves of Figs. 2 to 5.

The first portion of the curve of Fig. 2 shows the variation of Δ_1 with time at a very small strain amplitude, for a crystal which was mounted, without annealing, immediately after removal from the crucible. During the indicated time the oscillator remained isolated at the temperature 26°C. After 4 days the oscillator was removed from the apparatus, disassembled,

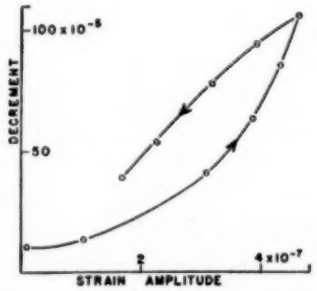


FIG. 4. The decrements measured when the strain amplitude is increased by steps are smaller than those when the strain amplitude is subsequently decreased.

and the zinc crystal was annealed for 1 hr. at 250°C. The remainder of the curve shows the variation of Δ_1 with time after reassembly and remounting.

Observations made with increasing strain amplitude cannot be repeated in reverse order (Fig. 4), nor can they be repeated at all until after an elapsed time which varies from several hours to several days. Curve *B* of Fig. 5 was obtained 4 hr. after curve *A*.

The curves of Fig. 6 show the variation of Δ_1 with maximum strain amplitude for an array of zinc crystals whose cylinder axes make various angles, θ, with the hexagonal axis of the crystal lattice. The procedure by which the data here plotted were obtained was as follows: After the crystals were removed from the molds and sawed to the proper length with a fine jeweler's saw, they were annealed for 2 hr. at 250°C in a vacuum, and then mounted. The measurements

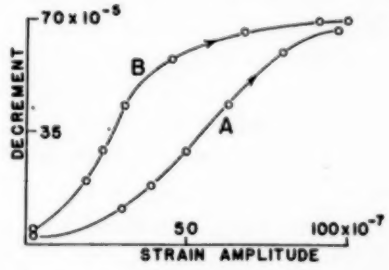


FIG. 5. Observations cannot be repeated until a considerable time has elapsed. The data for curve *B* were obtained 4 hours after those for curve *A*.

were made one or two days later, always after 10 minutes steady oscillation at each amplitude, and always on *increasing* amplitude. The vibration frequency for all crystals was about 38 kc per sec.

The decrement of the crystals is approximately inversely proportional to the vibration frequency, as is indicated by the typical set of data plotted in Fig. 7. These observations were made on the same crystal when vibrating, in the first instance, at 38 kc with one half-wave in the specimen, and in the second, at 76 kc with two half-waves in the specimen.

The associated variation of the resonance frequency, f_1, with maximum strain amplitude in the four crystals is shown in Fig. 8. A comparison of the data given in Figs. 6 and 8 reveals the significant fact that the ratio $r=\Delta_1 \div (\delta f_1/f_1)$ *is a constant independent of the strain amplitude.* The experimental confirmation of this is exhibited in Fig. 9, and the values of r for the different crystals are recorded in Table I.

The value of Young's modulus for an arbitrary direction in a hexagonal crystal is given in terms of the principal elastic moduli by the formula[12]

$$1/E = s_{11}\sin^4\theta + s_{33}\cos^4\theta + (2s_{13}+s_{44})\sin^2\theta\cos^2\theta.$$

[12] W. Voigt, *Lehrbuch der Kristallphysik*, p. 746.

Accordingly, the present measurements permit the evaluation of the principal adiabatic moduli s_{11} and s_{33}, and of the combination $(2s_{13}+s_{44})$. The corresponding isothermal quantities can be calculated with the aid of the thermal expansion coefficients measured by Gruneisen and Goens.[13] The results of the present measurements, together with those reported by other observers, are presented for comparison in Table II.

Before proceeding to a discussion of the foregoing data, which deal with the properties Δ_1 and δf_1 of the specimen *rod*, it is well to point out the way in which these can, if desired, be correlated with the properties of the specimen *material*. Such correlation, which is based on Eqs. (12) and (18), is greatly facilitated by the experimental close matching of the frequencies f_0 and f_1. In this event the variation of the local strain amplitude, U_1, along the rod is given very nearly by the formula

$$U_1 = U_1^0 \sin(\pi x/L_1), \tag{20}$$

and Eq. (12) may then be rewritten in the form

$$\Delta_1 = (2/L_1)\int_0^{L_1} \Delta \sin^2(\pi x/L_1)dx. \tag{21}$$

The variation of the decrement of the *material*, Δ, with *local* strain amplitude might now be obtained with the aid of Eqs. (20) and (21), by expanding Δ_1 in a power series in U_1^0, whose coefficients are determined by the curves of Fig. 6, and Δ in a power series in U_1, whose coefficients can then be obtained by equating the coefficients of like powers of U_1^0 in Eq. (21).

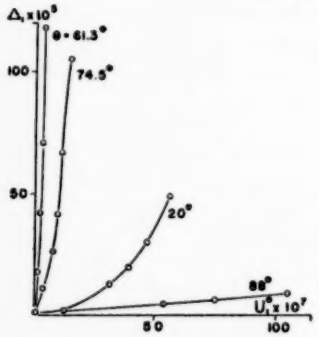


FIG. 6. Showing the variation of the decrement, Δ_1, with strain amplitude, U_1^0, in zinc crystals whose cylinder axes make various angles, θ, with the hexagonal axis.

[13] W. Voigt, reference 13. E. Gruneisen and E. Goens, Zeits. f. Physik **29**, 141 (1924).

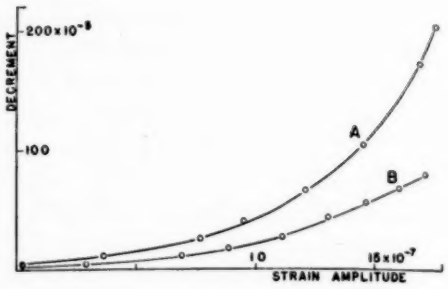


FIG. 7. The decrement is very nearly inversely proportional to the vibration frequency at all strain amplitudes. The vibration frequency for the data of curve *A* is 38 kc and of curve *B* is 76 kc.

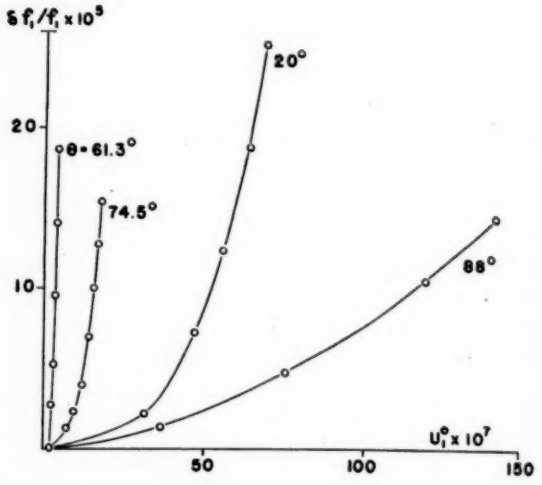


FIG. 8. Showing the fractional variation of the fundamental resonant frequency, $\delta f_1/f_1$, with the maximum strain amplitude, U_1^0, in zinc crystals whose cylinder axes make various angles, θ, with the hexagonal axis.

A similar procedure, applied to Eq. (18), would reveal the relation between E_1 and U_1.

A detailed calculation of the sort just outlined is not considered important in connection with the present data, but the similarity in form of Eqs. (18) and (21) is extremely significant. For this result, together with the observed mutual independence of the ratio r and U_1^0, implies that the relation between 2Δ and $\delta E_1/E_1$ is exactly the same as the relation between Δ_1 and $\delta f_1/f_1$, i.e., that

$$2\Delta \div (\delta E_1/E_1) = r. \tag{22}$$

DISCUSSION

The fact, revealed by data of the sort plotted in Fig. 7, that the decrement of single metal crystals varies inversely with the frequency of

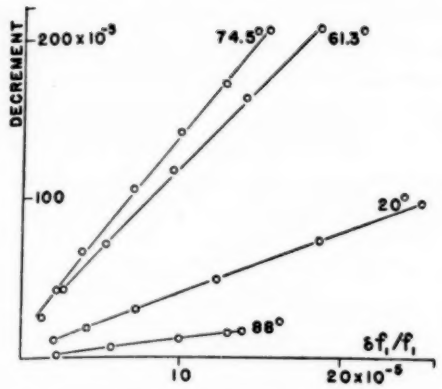


FIG. 9. Combination of the data plotted in Figs. 6 and 8 shows that the ratio of the decrement to the fractional change in resonant frequency is independent of the strain amplitude.

vibration suggests that the origin of the internal friction here studied is the same as that of the observed macroscopic plastic flow. The latter phenomenon has been ascribed by Taylor, Polanyi, Orowan and others to the propagation through the lattice of a specific type of "dislocation" in the atomic arrangement.[14] The nature of the process is illustrated in Fig. 10, which is similar to a figure in one of Taylor's papers. Fig. 10(A) represents a cross section through the lattice of a perfect crystal, the slip plane being supposed parallel to a horizontal row of atoms in

TABLE I. *Values of the ratio* $r=\Delta_1 \div (\delta f_1/f_1)$ *for different zinc crystals.*

$\theta°$	r
20.0	2.8
61.2	10.3
61.5	11.0
74.5	9.0
88.0	1.1

TABLE II. *The principal isothermal elastic moduli of crystalline zinc, as measured by various observers.*[1]

OBSERVER	MODULUS $\times 10^{13}$ CM²/DYNE		
	s_{11}	s_{33}	$2s_{13}+s_{44}$
Read	8.35	28.5	11.3
Goens	8.4	28.7	11.0
Tyndall and Hanson	7.70	27.66	10.54

[1] E. Goens, Ann. d. Physik **16**, 793 (1933); E. P. T. Tyndall, Phys. Rev. **47**, 398 (1935).

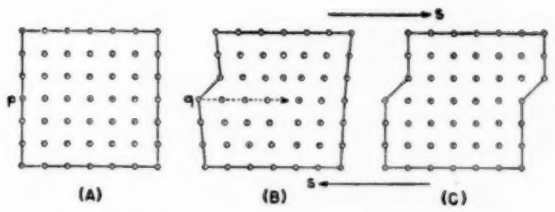


FIG. 10. Illustrating the formation and propagation of a dislocation (after Taylor).

the figure. A dislocation occurs when, in consequence of a weakening of the interatomic binding force, and atom, for example at p, "jumps" an atomic distance parallel to the slip plane into a new equilibrium position at q. The action of an applied shearing stress, S, on the atoms above and below the slip plane causes the dislocation thus formed to be propagated through the lattice in the manner indicated in Fig. 10(B). Thus the final configuration of the atomic group is as shown in Fig. 10(C).

The dislocations are assumed to originate at, and their propagation to be arrested by, microscopic flaws in the crystal due to occluded impurity, to surfaces of misfit, or to a mosaic superstructure. They occur spontaneously as a result of thermal agitation, and are produced by an applied stress. The propagation of each dislocation is accompanied by a dissipation of energy, associated with the atomic jumps, and a local slip to the amount of one atomic distance. The net contribution of such processes to the internal friction and inelastic strain depends on the number density of dislocations present in the crystal.

The mechanism here described is adequate to account qualitatively for the several phenomena revealed in Figs. 2 to 7. Thus, since the presence of the dislocations implies an internal potential energy in excess of that characteristic of the perfect lattice, the number density of dislocations decreases in time at a rate which depends on the material and the temperature (Fig. 2). Again, the number density present in a vibrating crystal is fixed when the rate of production of dislocations by the applied stress is equal to the natural rate of decrease, and the establishment of this equilibrium number requires a finite time (Fig. 3). Similarly, a finite time is required to establish the new equilibrium value

[14] G. I. Taylor, Proc. Roy. Soc. **A145**, 362, 388, 405 (1934); M. Polanyi, Zeits. f. Physik **89**, 660 (1934); E. Orowan, Zeits. f. Physik **89**, 605, 614, 634 (1934); **97**, 573 (1935); **98**, 382 (1936). For a complete bibliography see R. Houwink, *Elasticity, Plasticity, and Structure of Matter* (Cambridge, 1937).

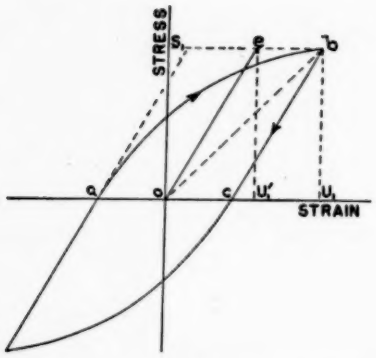


FIG. 11. Suggested form of the stress-strain diagram for single metal crystals.

associated with a decreased vibrational stress amplitude (Figs. 4 and 5). Lastly, the number density of dislocations increases with the amplitude of the component shearing stress parallel to the slip plane (Fig. 6). It remains to consider the variation of Young's modulus with the vibration amplitude.

The variation of E_1 with U_1, small though it is, nevertheless definitely indicates that the stress-strain relation is not that characteristic of *pure* plastic flow. A substance is plastic if, for a given strain, the stress decreases in time at a rate proportional to its own value. Thus the time variation of the stress in a plastic rod is given by the equation[15]

$$dS/dt = E(ds/dt) - pS, \tag{23}$$

where s is the strain and p is a constant which is, in the present experimental arrangement, related to Δ by the formula $p = 2f_1\Delta$. Hence the frequency variation of Δ is as stated at the beginning of this section. But it also follows from Eq. (23) that the stress-strain diagram is an ellipse, the slope of whose major axis, which is the effective value of Young's modulus, *does not vary with the strain amplitude.* It is evident, therefore, that the circumstances of pure plastic flow do not subsist over the entire cycle, and that the description offered by the ideal stress-strain ellipse must be modified.

It is suggested that plastic flow, associated with the propagation of dislocations, accompanies the *application* of a stress but not its *removal*, i.e., that the application of a stress produces what may properly be called "strain-hardening" in the material. The nature of the suggested stress-strain diagrams at very low and at larger amplitudes are indicated in Fig. 11. At very low amplitudes the strain is elastic, and the half-loop is the line *oe*. At larger amplitudes a plastic strain of amount $2(U_1 - U_1')$ takes place over the portion *ab* of the cycle, but, in consequence of the concurrent strain hardening, the portion *bc* is nearly elastic.

It must be remembered that Fig. 11 is not drawn to scale. Actually the plastic strain, shown by the intercept *ac*, is not larger than one-thousandth the strain amplitude, shown by the abscissa U_1. The stress-strain loop hugs its major axis so closely that the effective Young's modulus is always the slope of this axis. Accordingly, the value of E_1 at low amplitudes is S_1/U_1', and at larger amplitudes is S_1/U_1, and it follows that

$$\delta E_1/E_1 = ac/2U_1, \tag{24}$$

where δE_1 is the change in E_1 with amplitude.

Considerable support is lent the present hypothesis by the fact that it is not only in accord with the observed invariance of the ratio r (Eq. (22)) to change in U_1, but it leads to a value of this ratio which is in good agreement with observation. The area of the stress-strain diagram, which is the energy dissipated per unit volume per cycle, W^d, lies between $S_1 \cdot ac$ and $2S_1 \cdot ac$, and it follows immediately from Eqs. (1), (11), (22) and (24) that the value of r lies between 4 and 8, independently of U_1, and in agreement with the data of Table I. Larger values of r may be assumed to be the result of a state of affairs depicted by a downward flexure of the line *bc* in Fig. 11, and smaller values of that depicted by an upward flexure.

The origin of the variation of r with the angle θ, and the temperature dependence of the several phenomena, remain subjects for future investigation. In conclusion, the writer desires to acknowledge his indebtedness to the officers of the New Jersey Zinc Company, for their courtesy in supplying the extremely pure material from which the zinc crystals were grown; to Dr. W. F. Brown and Dr. Clarence Zener, for many helpful conversations; and to Dr. S. L. Quimby, for his guidance during the progress of the work.

[15] R. Houwink, reference 14, p. 7.

LETTERS TO THE EDITOR

Prompt publication of brief reports of important discoveries in physics may be secured by addressing them to this department. Closing dates for this department are, for the first issue of the month, the eighteenth of the preceding month, for the second issue, the third of the month. Because of the late closing dates for the section no proof can be shown to authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents.

Communications should not in general exceed 600 words in length.

The Effects of Order and Disorder on Secondary Electron Emission

In previous papers[1, 2] it has been pointed out that an interaction between a primary electron and a lattice electron that results in secondary electron emission takes place in such a way as to transfer the lattice electron into the level in a higher band for which the reduced wave vector is the same as that of the originally occupied state. Order in an alloy produces a long period regularity in the lattice fields which should result in splitting some of the bands and making possible transitions between levels that could not exhibit interaction processes in the disordered alloy. Consequently, one might expect the order-disorder transition in an alloy to be accompanied by changes in the secondary electron emission properties. The magnitude of this effect cannot be predicted easily; it should depend largely on the difference between the lattice fields produced by the types of atom involved.

An attempt has been made to observe the effect of order on the secondary electron emission from a Cu_3Au target. The experimental apparatus and the technique were the same as those employed in earlier temperature studies on nickel, cobalt and molybdenum.[2] Measurements were preceded by baking of the tube, glowing of all the metal parts and a treatment of the target which included a few minutes of heating at 700°C, followed by 10 or 20 hours at 400–450°C. The pressure was $<10^{-7}$ mm Hg during the ensuing experiments. Figure 1 is a typical result. The ordinate, V_1, is the bombarding voltage for which the secondary electron emission coefficient is exactly unity. These measurements were made for ascending temperatures, with a heating rate of 45 degrees per hour. The measurements were preceded by a 10-hour nearly uniform cooling from the annealing temperature of 450°C.

This alloy begins to disorder at around 250°C,[3] while most of the disordering occurs above 350 or 360°C. Complete disorder is believed to exist above 391°C.[4] The curve of Fig. 1 gives a suggestion of some such behavior, but is not at all conclusive. The experiment has been repeated a number of times under a variety of conditions. In particular, the results are not appreciably altered by a preliminary 12-hour heating at 360°C to ensure completeness of the initial ordering of the target.

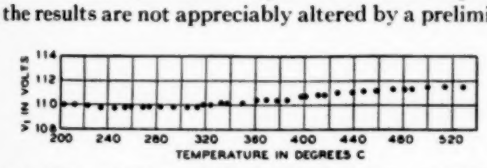


FIG. 1. Effect of order on the secondary electron emission from Cu_3Au. V_1 = bombarding voltage for which secondary electron emission coefficient is unity.

From this work it can only be concluded that the secondary emission of an ordered specimen of Cu_3Au differs from that of the same specimen in a disordered state by not more than one or two percent.

D. E. WOOLDRIDGE
C. D. HARTMAN

Bell Telephone Laboratories,
New York, New York,
April 24, 1940.

[1] D. E. Wooldridge, Phys. Rev. **56**, 562 (1939).
[2] D. E. Wooldridge, Phys. Rev., this issue.
[3] F. C. Nix and D. MacNair, Phys. Rev. **55**, 605 (1939).
[4] C. Sykes and F. W. Jones, Proc. Roy. Soc. **157**, 213 (1936).

The Zeeman Effect in Krypton and Xenon

Measurements of the Zeeman effect of about 150 lines in the spectrum of Kr I and about the same number in X I have elicited several interesting points with regard to these spectra.

In the case of the lowest p^5s configuration of neon, argon, and krypton, the agreement between experimental and theoretical g values is practically perfect; but the lowest p^5s level of xenon shows disagreement with theory as well as a violation of the g sum rule.

The lowest p^5p configurations of neon, argon, and krypton while not yielding experimental g values in agreement with least-square calculations, still satisfy the g sum rule for $J=1$. This is not the case for xenon. Our results are not in agreement with those of Pogány.[1]

Both of the discrepancies between theory and experiment may be attributed to configuration interaction. There are several other examples of this phenomenon and for Kr $4p^56p$ the disturbing factor is shown to be the effect of $4p^55f$.

The $4p^54f$ configuration in Kr shows evidence that the levels of this type of configuration are double, a fact predicted by Shortley and Fried.[2]

In addition, several changes of assignment of levels are suggested by the Zeeman patterns. Complete results will be published soon.

J. B. GREEN
D. W. BOWMAN
E. H. HURLBURT

Mendenhall Laboratory of Physics,
Ohio State University,
Columbus, Ohio,
July 19, 1940.

[1] B. Pogány, Zeits. f. Physik **93**, 364 (1935).
[2] G. H. Shortley and B. Fried, Phys. Rev. **54**, 749 (1938).

Photographic Plate Spectrum of *d–d* Neutrons

Recently in this laboratory[1] a search was made for a low energy group of neutrons arising from the *d-d* reaction by an observation of helium recoils produced in a cloud chamber. No indication of such a group was found. The fact that the short recoils previously observed in methane[2] and in hydrogen[3] were of the range one would expect for deuterons formed by proton capture of the high energy *d–d* neutrons would explain the experimental facts, but the ratio of the probability of capture to the probability of collision for neutrons of 2.5 Mev energy in hydrogen is theoretically much too small to account for the relative intensities of the two groups on this basis.

In the present work we thought to test the idea of radiative capture of neutrons by using the technique of photographic plates,[4] in which a deuteron could be distinguished from a proton. Also, photographic plates, compared to heavier detecting equipment, give less scattering, and hence we could study the possibility that the low energy group previously reported was due to this cause.

Magnetically analyzed 700-kv deuterons, accelerated by the Rice pressure Van de Graaff generator, fell on a heavy paraffin target of about 20 kv thickness for one microampere-hour. Photographic plates were placed 12 cm from the target at 0° and at 90° to the bombarding beam. Following the technique which one of us (HTR) has previously used,[5] we have analyzed the two plates for proton recoils. Only those recoils within 12° of the forward direction were tabulated; a photomicrograph of a recoil proton track is shown in Fig. 1.

The range-number curve for each plate was converted to a range-energy curve, shown in Fig. 2, by calculating the stopping power of the photographic emulsion. With $Q=3.32$ Mev, one calculates the energy of the neutrons emitted at 90° to the 700-kv bombarding beam to be 2.67 Mev, while neutrons emitted at 0° have an energy of 3.82 Mev. These groups showed mean ranges of 73 and 138 microns, respectively. Hence the stopping power of the emulsion is 6.4 microns per cm of air in the first case, and 6.6 microns per cm of air in the second. Some α-tracks, caused by radioactive contamination of the emulsion, were also observed. These, however, show a much greater density and are thus easy to distinguish from the tracks produced by proton recoils.

It will be observed that there is no evidence of a short range group of recoils in the data at 90°. This shows that radiative capture cross section is indeed very small, and also shows that there is no low energy group of neutrons from the *d–d* reaction. The data at 0° confirm these conclusions, since the weak group at 2.5 Mev is thought to have arisen from bombardment of deuterium contamination on a wall of the magnet box by the molecular beam. Since the molecular beam is deviated through only a small angle before striking the wall, neutrons from this source must be ejected at nearly 90° to the molecular beam in order to strike the photographic plates. The tail present beyond the maximum on the curve obtained from the plate at 90° to the bombarding beam is doubtless due to the scattering by the target holder of high energy neutrons ejected at smaller angles.

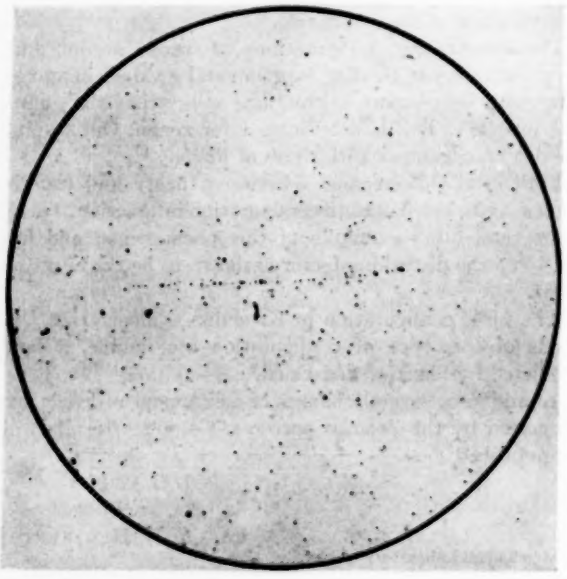


FIG. 1. A photomicrograph (750 X) of a 78-micron proton recoil in a photographic emulsion. The proton in this case proceeded from left to right. Increased linear grain density and small angle deflection are characteristic of the termination of such tracks.

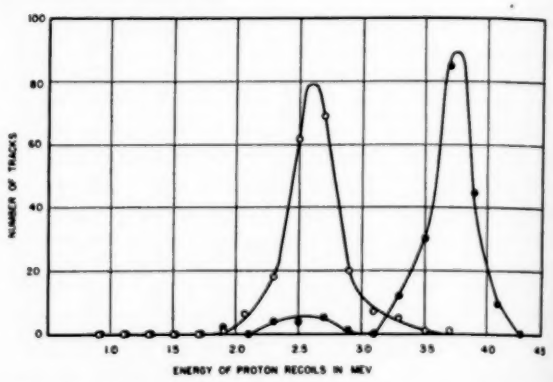


FIG. 2. The *d–d* neutron spectrum as observed in photographic plates placed at 90° (open circles) and at 0° (dots) to a 700-kv bombarding beam.

We therefore conclude that the radiative capture cross section is at least 100 times smaller than the scattering cross section for 2.5-Mev neutrons. We further conclude that there is only a single group of neutrons from the *d–d* reaction. The short range groups previously reported could have arisen through scattering or through contamination of the cloud chamber by some radioactive substance which could yield a radioactive gas.

HUGH T. RICHARDS
EMMETT HUDSPETH

The Rice Institute,
Houston, Texas,
July 20, 1940.

[1] E. Hudspeth and H. Dunlap, Phys. Rev. **57**, 971 (1940).
[2] T. W. Bonner, Phys. Rev. **52**, 685 (1937) and **53**, 711 (1938).
[3] E. Hudspeth and H. Dunlap, Phys. Rev. **55**, 587 (1939).
[4] Ilford Special Halftone plates, of 100 microns thickness, were used in this work.
[5] An account of this investigation (of the Li neutron spectrum) will be published soon and will include details of the particular technique employed.

The Beta-Radiation from Si^{27} and P^{30}

In continuation of our work on the light nuclei we have studied the upper limit of the positrons from Si^{27} and P^{30}. From the momentum spectra obtained in an alcohol-air-filled cloud chamber we obtain upper limits of 3.54 ± 0.1 Mev and 3.0 ± 0.1 Mev, respectively.

Si^{27} was formed by bombarding Al^{27} with 6.4-Mev protons, the reaction being $Al^{27}(p, n)Si^{27}$. In order to get the targets into the cloud chamber before this 4.9-second activity decayed, it was necessary to do the bombarding outside the main cyclotron vacuum chamber. It was also found advantageous to keep an atmosphere of hydrogen around the target during bombardment to avoid possible air contamination. The cloud chamber was set up near the cyclotron, but far enough away to escape the leakage magnetic field. Rather than attempt the difficult task of shielding the cloud chamber from the neutron and gamma-field of the cyclotron, we synchronized the ion-source filament heating current with the cloud chamber and in this way turned off the beam before every expansion. Control tests showed that under these conditions the radiations from the cyclotron had no visible effect on the cloud chamber. Magnetic field measurements over the cloud-chamber volume were made in several ways, all of which were consistent to within one percent. Our most reliable method was to place the evacuated plate holder of a photographic beta-ray spectrograph in the middle of the cloud-chamber Helmholtz coils. Since there are many well-known internal conversion lines in the active deposit of thorium they were used as calibration checks on the field.

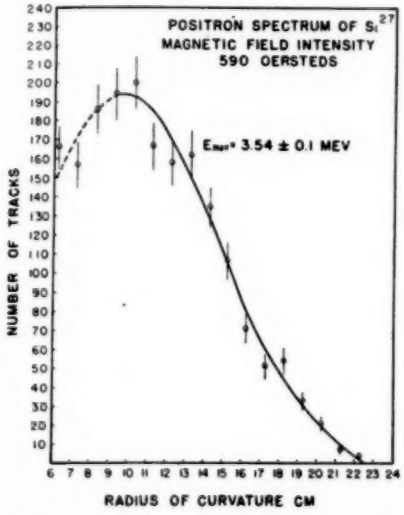


FIG. 1. Momentum spectrum of Si^{27}. Cloud-chamber window correction of 0.03 Mev has been added to the visual upper limit of both Si^{27} and P^{30}.

Although the disintegration energy of Si^{27} quoted above is slightly higher than the 3.4 Mev expected[1] on the basis of a nuclear volume proportional to the mass number, we prefer to attribute the discrepancy to errors for which no reliable estimate is possible. Scattering in the gas of the cloud chamber would be the most likely source of error although every effort was made to reject scattered tracks. Our value may be compared with the still higher value of 3.74 Mev which has already been obtained at Rochester.[2]

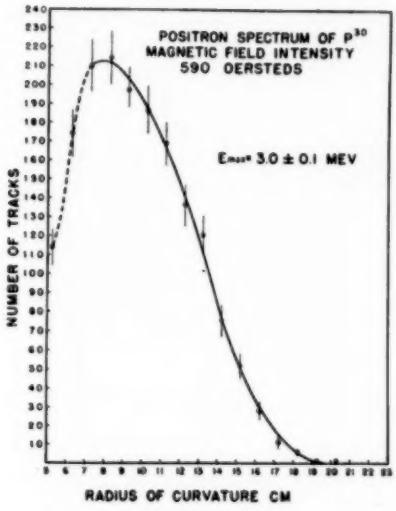


FIG. 2. Momentum spectrum of P^{30}.

In Fig. 1 is shown the $H\rho$ distribution of approximately 2000 positron tracks from Si^{27}. The dotted portion of the curve indicates the region below which no precision was attempted since only the upper limit was under scrutiny. Experimental errors are indicated by vertical lines of length determined by taking the square root of the number of tracks.

P^{30} was investigated because so many discordant measurements are quoted in the literature. We formed this isotope by the reaction $Si^{30}(p, n)P^{30}$, identification being made through the half-life of 2.5 minutes. Figure 2 shows the $H\rho$ diagram of the emitted positron. Our value agrees within the limits of error with that of Curie and Joliot.[3] An accurate knowledge of P^{30} is of particular interest since it belongs to the nuclear type $4K+2$ and should therefore give information on spin dependent forces.[4]

W. H. BARKAS
E. C. CREUTZ
L. A. DELSASSO
R. B. SUTTON
M. G. WHITE

Palmer Physical Laboratory,
Princeton University,
Princeton, New Jersey,
July 17, 1940.

[1] White, Delsasso, Fox and Creutz, Phys. Rev. **56**, 512 (1939).
[2] McCreary, Kuerti and Van Voorhis, Phys. Rev. **57**, 351 (1940).
[3] I. Curie and F. Joliot, J. de phys. et rad. **5**, 153 (1934).
[4] E. Wigner, Phys. Rev. **56**, 519 (1939).